Econometrica Supplementary Material

SUPPLEMENT TO “INFERENCE BASED ON CONDITIONAL MOMENT INEQUALITIES”

(*Econometrica*, Vol. 81, No. 2, March 2013, 609–666)

BY DONALD W. K. ANDREWS AND XIAOXIA SHI

CONTENTS

 DOI: 10.3982/ECTA9370

11. OUTLINE

THIS SUPPLEMENT INCLUDES six appendices.

Supplemental Appendix A gives proofs of Theorems 1 and 2(a).

Supplemental Appendix B provides a number of supplemental results to the main paper. These include:

(i) results for Kolmogorov–Smirnov (KS) and approximate Cramér–von Mises (A-CvM) tests and CS's in Section 13.1,

(ii) three additional examples of collections $\mathcal{G}$ and probability measures Q that satisfy Assumptions CI, M, FA(e), and Q in Section 13.2,

(iii) the verification of Assumption GMS2(a) under some conditions on S, Q, and α in Section 13.3,

(iv) an illustration of the verification of Assumptions LA1–LA3 in Section 13.4,

(v) an illustration of some uniformity issues that arise with infinite-dimensional nuisance parameters in Section 13.5,

(vi) an illustration of problems with pointwise asymptotics in Section 13.6, and

(vii) coverage probability results for subsampling tests and CS's under drifting sequences of distributions in Section 13.7.

Supplemental Appendix C provides proofs of the results that are stated in the main paper but are not proved in Supplemental Appendix A. These include:

(i) the proofs of Lemmas 2 and 3 and Theorem 2(b) in Section 14.1,

(ii) the proofs of Lemma 4 and Theorem 3 concerning fixed alternatives in Section 14.2,

(iii) the proof of Theorem 4 concerning local power in Section 14.3, and

(iv) the proof of Lemma 1 concerning the verification of Assumptions S1–S4 in Section 14.4.

Supplemental Appendix D provides proofs of the results stated in Supplemental Appendix B. These include:

(i) the proofs of Kolmogorov–Smirnov and approximate Cramér–von Mises results in Section 15.1,

(ii) the proof of Lemma B2 in Section 15.2,
(iii) the proofs of Theorems B4 and B5 regarding uniformity issues in Section 15.3, and
(iv) the proofs of the subsampling results in Section 15.4.
Supplemental Appendix E proves Lemma A1, which is stated in Supplemental Appendix A.

Supplemental Appendix F provides the simulation results for the mean selection and interval-outcome regression models and additional material (and results) concerning the simulations in the quantile selection and entry game models.

12. SUPPLEMENTAL APPENDIX A

This appendix provides proofs of the uniform asymptotic coverage probability results for GMS and PA CS's. In particular, it proves Theorems 1 and 2(a). Proofs of the other results stated in the paper are given in Supplemental Appendix C.

12.1. *Proof of Theorem 1*

The following lemma is used in the proofs of Theorems 1, 2, 3, and 4. It establishes a functional CLT and uniform LLN for certain independent non-identically distributed empirical processes.

Let h_2 denote a $k \times k$-matrix-valued covariance kernel on $\mathcal{G} \times \mathcal{G}$ (such as an element of $\mathcal{H}_2$).

DEFINITION SubSeq(h_2): SubSeq(h_2) is the set of subsequences $\{(\theta_{a_n}, F_{a_n}): n \geq 1\}$, where $\{a_n : n \geq 1\}$ is some subsequence of $\{n\}$, for which

$$\text{(i)} \quad \lim_{n\to\infty} \sup_{g,g^*\in\mathcal{G}} \left\| h_{2,F_{a_n}}(\theta_{a_n}, g, g^*) - h_2(g, g^*) \right\| = 0,$$

(ii) $\theta_{a_n} \in \Theta$, (iii) $\{W_i : i \geq 1\}$ are i.i.d. under F_{a_n}, (iv) $\mathrm{Var}_{F_{a_n}}(m_j(W_i, \theta_{a_n})) > 0$ for $j = 1, \ldots, k$, for $n \geq 1$, (v) $\sup_{n\geq 1} E_{F_{a_n}} |m_j(W_i, \theta_{a_n})/\sigma_{F_{a_n},j}(\theta_{a_n})|^{2+\delta} < \infty$ for $j = 1, \ldots, k$, for some $\delta > 0$, and (vi) Assumption M holds with F_{a_n} in place of F and F_n in Assumptions M(b) and M(c), respectively.

The sample paths of the Gaussian process $\nu_{h_2}(\cdot)$, which is defined in (4.2) and appears in the following lemma, are bounded and uniformly ρ-continuous a.s. The pseudo-metric ρ on $\mathcal{G}$ is a pseudo-metric commonly used in the empirical process literature:

$$\rho^2(g, g^*) = \mathrm{tr}\big(h_2(g, g) - h_2(g, g^*) - h_2(g^*, g) + h_2(g^*, g^*)\big). \tag{12.1}$$

For $h_2(\cdot,\cdot) = h_{2,F}(\theta,\cdot,\cdot)$, where $(\theta,F)\in\mathcal{F}$, this metric can be written equivalently as

$$\begin{aligned}(12.2)\qquad &\rho^2(g,g^*) = E_F\left\|D_F^{-1/2}(\theta)\left[\widetilde{m}(W_i,\theta,g)-\widetilde{m}(W_i,\theta,g^*)\right]\right\|^2, \quad \text{where}\\ &\widetilde{m}(W_i,\theta,g) = m(W_i,\theta,g) - E_F m(W_i,\theta,g).\end{aligned}$$

LEMMA A1: *For any subsequence* $\{(\theta_{a_n},F_{a_n}):n\geq 1\}\in \mathrm{SubSeq}(h_2)$,
(a) $\nu_{a_n,F_{a_n}}(\theta_{a_n},\cdot)\Rightarrow \nu_{h_2}(\cdot)$ *as* $n\to\infty$ (*as processes indexed by* $g\in\mathcal{G}$), *and*
(b) $\sup_{g,g^*\in\mathcal{G}}\|\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n},g,g^*)-h_2(g,g^*)\|\to_p 0$ *as* $n\to\infty$.

COMMENTS: (i) The proof of Lemma A1 is given in Supplemental Appendix E. Part (a) is proved by establishing the manageability of $\{m(W_i,\theta_{a_n},g) - E_{F_{a_n}}m(W_i,\theta_{a_n},g):g\in\mathcal{G}\}$ and by establishing a functional CLT for R^k-valued i.n.i.d. empirical processes with the pseudo-metric ρ by using the functional CLT in Pollard (1990, Thm. 10.2) for real-valued empirical processes. Part (b) is proved using a maximal inequality given in Pollard (1990, (7.10)).

(ii) To obtain uniform asymptotic coverage probability results for CS's, Lemma A1 is applied with $(\theta_{a_n},F_{a_n})\in\mathcal{F}$ for all $n\geq 1$ and $h_2\in\mathcal{H}_2$. In this case, conditions (ii)–(vi) in the definition of $\mathrm{SubSeq}(h_2)$ hold automatically by the definition of $\mathcal{F}$. To obtain power results under fixed and local alternatives, Lemma A1 is applied with $(\theta_{a_n},F_{a_n})\notin\mathcal{F}$ for all $n\geq 1$ and h_2 may or may not be in $\mathcal{H}_2$.

PROOF OF THEOREM 1 : First, we prove part (a). Let $\{(\theta_n,F_n)\in\mathcal{F}:n\geq 1\}$ be a sequence for which $h_{2,F_n}(\theta_n)\in\mathcal{H}_{2,\mathrm{cpt}}$ for all $n\geq 1$ and the term in square brackets in Theorem 1(a), evaluated at (θ_n,F_n), differs from its supremum over $(\theta,F)\in\mathcal{F}$ with $h_{2,F}(\theta)\in\mathcal{H}_{2,\mathrm{cpt}}$ by δ_n or less, where $0<\delta_n\to 0$ as $n\to\infty$. Such a sequence always exists. To prove part (a), it suffices to show that part (a) holds with the supremum deleted and with (θ,F) replaced by (θ_n,F_n).

By the compactness of $\mathcal{H}_{2,\mathrm{cpt}}$, given any subsequence $\{u_n:n\geq 1\}$ of $\{n\}$, there exists a subsubsequence $\{a_n:n\geq 1\}$ for which $d(h_{2,F_{a_n}}(\theta_{a_n}),h_{2,0})\to 0$ as $n\to\infty$ for some $\theta_0\in\Theta$, where d is defined in (5.6), and some $h_{2,0}\in\mathcal{H}_{2,\mathrm{cpt}}$. This and $(\theta_{a_n},F_{a_n})\in\mathcal{F}$ for all $n\geq 1$ imply that $\{(\theta_{a_n},F_{a_n}):n\geq 1\}\in\mathrm{SubSeq}(h_{2,0})$.

Now, by Lemma A1, we have

$$(12.3)\qquad \begin{pmatrix}\nu_{a_n,F_{a_n}}(\theta_{a_n},\cdot)\\ \widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n},\cdot)\end{pmatrix}\Rightarrow\begin{pmatrix}\nu_{h_{2,0}}(\cdot)\\ h_{2,0}(\cdot)\end{pmatrix}\quad\text{as}\quad n\to\infty$$

as stochastic processes on $\mathcal{G}$, where $\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n},g)=\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n},g,g)$ and $h_{2,0}(g)=h_{2,0}(g,g)$.

Given this, by the almost sure representation theorem (e.g., see Pollard (1990, Thm. 9.4)), there exists a probability space and random quantities $\tilde{\nu}_{a_n}(\cdot)$,

$\tilde{h}_{2,a_n}(\cdot)$, $\tilde{\nu}_0(\cdot)$, and $\tilde{h}_2(\cdot)$ defined on it such that (i) $(\tilde{\nu}_{a_n}(\cdot), \tilde{h}_{2,a_n}(\cdot))$ has the same distribution as $(\nu_{a_n,F_{a_n}}(\theta_{a_n},\cdot), \widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n},\cdot))$, (ii) $(\tilde{\nu}_0(\cdot), \tilde{h}_2(\cdot))$ has the same distribution as $(\nu_{h_{2,0}}(\cdot), h_{2,0}(\cdot))$, and

$$\text{(12.4)}\qquad \text{(iii)}\quad \sup_{g\in\mathcal{G}}\left\|\begin{pmatrix}\tilde{\nu}_{a_n}(g)\\ \tilde{h}_{2,a_n}(g)\end{pmatrix} - \begin{pmatrix}\tilde{\nu}_0(g)\\ \tilde{h}_2(g)\end{pmatrix}\right\| \to 0 \quad \text{as} \quad n\to\infty \quad \text{a.s.}$$

Because $h_{2,0}(\cdot)$ is deterministic, condition (ii) implies that $\tilde{h}_2(\cdot) = h_{2,0}(\cdot)$ a.s.

Define

$$\text{(12.5)}\qquad \begin{aligned} \tilde{h}^{\varepsilon}_{2,a_n}(\cdot) &= \tilde{h}_{2,a_n}(\cdot) + \varepsilon\cdot \mathrm{Diag}\big(\tilde{h}_{2,a_n}(1_k)\big),\\ \tilde{T}_{a_n} &= \int S\big(\tilde{\nu}_{a_n}(g) + h_{1,a_n,F_{a_n}}(\theta_{a_n},g), \tilde{h}^{\varepsilon}_{2,a_n}(g)\big)\,dQ(g),\\ h^{\varepsilon}_{2,0}(\cdot) &= h_{2,0}(\cdot) + \varepsilon I_k,\\ \tilde{T}_{a_n,0} &= \int S\big(\tilde{\nu}_0(g) + h_{1,a_n,F_{a_n}}(\theta_{a_n},g), h^{\varepsilon}_{2,0}(g)\big)\,dQ(g). \end{aligned}$$

By construction, $\tilde{T}_{a_n}$ and $T_{a_n}(\theta_{a_n})$ have the same distribution, and $\tilde{T}_{a_n,0}$ and $T(h_{a_n,F_{a_n}}(\theta_{a_n}))$ have the same distribution for all $n \geq 1$.

Hence, to prove part (a), it suffices to show that

$$\text{(12.6)}\qquad A = \limsup_{n\to\infty}\big[P_{F_{a_n}}(\tilde{T}_{a_n} > x_{h_{a_n,F_{a_n}}(\theta_{a_n})}) - P(\tilde{T}_{a_n,0} + \delta > x_{h_{a_n,F_{a_n}}(\theta_{a_n})})\big] \leq 0.$$

Below we show that

$$\text{(12.7)}\qquad \tilde{T}_{a_n} - \tilde{T}_{a_n,0} \to 0 \quad \text{as} \quad n\to\infty \quad \text{a.s.}$$

Let

$$\text{(12.8)}\qquad \begin{aligned} \widetilde{\Delta}_n &= 1\big(\tilde{T}_{a_n,0} + (\tilde{T}_{a_n} - \tilde{T}_{a_n,0}) > x_{h_{a_n,F_{a_n}}(\theta_{a_n})}\big) - 1(\tilde{T}_{a_n,0} + \delta > x_{h_{a_n,F_{a_n}}(\theta_{a_n})})\\ &= \widetilde{\Delta}^+_n - \widetilde{\Delta}^-_n, \quad \text{where}\\ \widetilde{\Delta}^+_n &= \max\{\widetilde{\Delta}_n, 0\} \in [0,1] \quad \text{and}\\ \widetilde{\Delta}^-_n &= \max\{-\widetilde{\Delta}_n, 0\} \in [0,1]. \end{aligned}$$

By (12.7) and $\delta > 0$, $\lim_{n\to\infty}\widetilde{\Delta}^+_n = 0$ a.s. Hence, by the bounded convergence theorem,

$$\text{(12.9)}\qquad \begin{aligned} &\lim_{n\to\infty} E_{F_{a_n}}\widetilde{\Delta}^+_n = 0,\\ &A = \limsup_{n\to\infty} E_{F_{a_n}}\widetilde{\Delta}_n \end{aligned}$$

$$= \limsup_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}_n^+ - \liminf_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}_n^-$$
$$= -\liminf_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}_n^- \le 0.$$

Hence, (12.6) holds and the proof of part (a) is complete, except for (12.7).

To prove part (b), analogous results to (12.6), (12.8), and (12.9) hold by analogous arguments.

It remains to show (12.7). We do so by fixing a sample path ω and using the bounded convergence theorem (because $\tilde{T}_{a_n}$ and $\tilde{T}_{a_n,0}$ are both integrals over $g \in \mathcal{G}$ with respect to the measure Q). Let $\tilde{\Omega}$ be the collection of all $\omega \in \Omega$ such that $(\tilde{\nu}_{a_n}(g), \tilde{h}_{2,a_n}(g))(\omega)$ converges to $(\tilde{\nu}_0(g), h_{2,0}(g))(\omega)$ uniformly over $g \in \mathcal{G}$ as $n \to \infty$ and $\sup_{g\in\mathcal{G}} \|\tilde{\nu}_0(g)(\omega)\| < \infty$. By (12.4) and $\tilde{h}_2(\cdot) = h_{2,0}(\cdot)$ a.s., $P(\tilde{\Omega}) = 1$. Consider a fixed $\omega \in \tilde{\Omega}$. By Assumption S2 and (12.4), for all $g \in \mathcal{G}$,

$$\sup_{\mu\in[0,\infty)^p\times\{0\}^v} \big|S\big(\tilde{\nu}_{a_n}(g)(\omega) + \mu, \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)\big) - S\big(\tilde{\nu}_0(g)(\omega) + \mu, h^{\varepsilon}_{2,0}(g)\big)\big| \to 0 \tag{12.10}$$

as $n \to \infty$. Thus, for all $g \in \mathcal{G}$ and all $\omega \in \widetilde{\Omega}$,

$$S\big(\tilde{\nu}_{a_n}(g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)\big) - S\big(\tilde{\nu}_0(g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), h^{\varepsilon}_{2,0}(g)\big) \to 0 \quad \text{as} \quad n \to \infty. \tag{12.11}$$

Next, we show that, for fixed $\omega \in \tilde{\Omega}$, the first summand on the left-hand side of (12.11) is bounded by a constant. Let $0 < \chi < 1$. By (12.4), there exists $N < \infty$ such that, for all $n \ge N$,

$$\sup_{g\in\mathcal{G}} \big\|\tilde{\nu}_{a_n}(g)(\omega) - \tilde{\nu}_0(g)(\omega)\big\| < \chi \quad \text{and} \tag{12.12}$$
$$\big\|\mathrm{Diag}\big(\tilde{h}_{2,a_n}(1_k)\big)(\omega) - I_k\big\| < \chi,$$

using the fact that $\mathrm{Diag}(h_{2,0}(1_k)) = I_k$ by construction. Let $B_\chi(\omega) = \sup_{g\in\mathcal{G}} \|\tilde{\nu}_0(g)(\omega)\| + \chi$. Then, for all $n \ge N$,

$$\sup_{g\in\mathcal{G}} \big\|\tilde{\nu}_{a_n}(g)(\omega)\big\| \le B_\chi(\omega) < \infty. \tag{12.13}$$

First, consider the case where no moment equalities are present, that is, $v = 0$ and $k = p$. In this case, for $n \ge N$, we have: for all $g \in \mathcal{G}$,

$$0 \le S\big(\tilde{\nu}_{a_n}(g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)\big) \tag{12.14}$$
$$\le S\big(\tilde{\nu}_{a_n}(g)(\omega), \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)\big)$$

$$\leq S\big(-B_\chi(\omega)1_p, \varepsilon \cdot \mathrm{Diag}\big(\tilde{h}_{2,a_n}(1_p)\big)\big)$$
$$\leq S\big(-B_\chi(\omega)1_p, \varepsilon(1-\chi)I_p\big),$$

where the first inequality holds by Assumption S1(c), the second inequality holds by Assumption S1(b) and $h_{1,a_n,F_{a_n}}(\theta_{a_n}, g) \geq 0_p$ (which holds because $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$), the third inequality holds by Assumption S1(b) and (12.13) as well as by Assumption S1(e) and the definition of $\tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)$ in (12.5), and the last inequality holds by Assumption S1(e) and (12.12). For fixed $\omega \in \tilde{\Omega}$, the constant $S(-B_\chi(\omega)1_p, \varepsilon(1-\chi)I_p)$ bounds the first summand on the left-hand side of (12.11) for all $n \geq N$.

For the case where $v > 0$, the third inequality in (12.14) needs to be altered because $S(m, \Sigma)$ is not assumed to be non-increasing in m_{II}, where $m = (m_{\mathrm{I}}', m_{\mathrm{II}}')'$. In this case, for the bound with respect to the last v elements of $\tilde{\nu}_{a_n}(g)(\omega)$, denoted by $\tilde{\nu}_{a_n,\mathrm{II}}(g)(\omega)$, we use the continuity condition on $S(m, \Sigma)$, that is, Assumption S1(d), which yields uniform continuity of $S(-B_\chi(\omega)1_p, m_{\mathrm{II}}, \varepsilon(1-\chi)I_k)$ over the compact set $\{m_{\mathrm{II}} : \|m_{\mathrm{II}}\| \leq B_\chi(\omega) < \infty\}$ and delivers a finite bound because $\sup_{g \in \mathcal{G}, n \geq 1} \|\tilde{\nu}_{a_n,\mathrm{II}}(g)(\omega)\| \leq B_\chi(\omega)$.

By an analogous but simpler argument, for fixed $\omega \in \tilde{\Omega}$, the second summand on the left-hand side of (12.11) is bounded by a constant.

Hence, the conditions of the bounded convergence theorem hold, and for fixed $\omega \in \tilde{\Omega}$, $\tilde{T}_{a_n}(\omega) - \tilde{T}_{a_n,0}(\omega) \to 0$ as $n \to \infty$. Thus, (12.7) holds and the proof is complete. *Q.E.D.*

12.2. *Proof of Theorem 2(a)*

For GMS CS's, Theorem 2(a) follows immediately from the following three lemmas. The PA critical value is a GMS critical value with $\varphi_n(x) = 0$ for all $x \in R$ and this function $\varphi_n(x)$ satisfies Assumption GMS1 (though not Assumption GMS2(b)). Hence, Theorem 2(a) for GMS CS's covers PA CS's.

LEMMA A2: *Suppose Assumptions* M, S1, *and* S2 *hold*. *Then*, *for every compact subset* $\mathcal{H}_{2,\mathrm{cpt}}$ *of* $\mathcal{H}_2$ *and all* $\delta > 0$,

$$\limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,\mathrm{cpt}}}} P_F\big(T_n(\theta) > c_0\big(h_{n,F}(\theta), 1-\alpha\big) + \delta\big) \leq \alpha.$$

LEMMA A3: *Suppose Assumptions* M, S1, *and* GMS1 *hold*. *Then*, *for every compact subset* $\mathcal{H}_{2,\mathrm{cpt}}$ *of* $\mathcal{H}_2$,

$$\lim_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,\mathrm{cpt}}}} P_F\big(c\big(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha\big)$$
$$< c\big(h_{1,n,F}(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha\big)\big) = 0.$$

LEMMA A4: *Suppose Assumptions* M, S1, *and* S2 *hold*. *Then*, *for every compact subset* $\mathcal{H}_{2,\text{cpt}}$ *of* $\mathcal{H}_2$ *and for all* $0 < \delta < \eta$ (*where* η *is as in the definition of* $c(h, 1-\alpha)$),

$$\lim_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,\text{cpt}}}} P_F\big(c\big(h_{1,n,F}(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha\big) < c_0\big(h_{1,n,F}(\theta), h_{2,F}(\theta), 1-\alpha\big) + \delta\big) = 0.$$

The following lemma is used in the proof of Lemma A4.

LEMMA A5: *Suppose Assumptions* M, S1, *and* S2 *hold*. *Let* $\{h_{2,n} : n \geq 1\}$ *and* $\{h^*_{2,n} : n \geq 1\}$ *be any two sequences of* $k \times k$*-valued covariance kernels on* $\mathcal{G} \times \mathcal{G}$ *such that* $d(h_{2,n}, h^*_{2,n}) \to 0$ *and* $d(h_{2,n}, h_{2,0}) \to 0$ *for some* $k \times k$*-valued covariance kernel* $h_{2,0}$ *on* $\mathcal{G} \times \mathcal{G}$. *Then*, *for all* $\eta_1 > 0$ *and all* $\delta > 0$,

$$\liminf_{n\to\infty} \inf_{h_1\in\mathcal{H}_1} \big[c_0(h_1, h_{2,n}, 1-\alpha+\eta_1) + \delta - c_0\big(h_1, h^*_{2,n}, 1-\alpha\big)\big] \geq 0.$$

PROOF OF LEMMA A2: For all $\delta > 0$, we have

$$\begin{aligned}(12.15)\quad & \limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,\text{cpt}}}} P_F\big(T_n(\theta) > c_0\big(h_{n,F}(\theta), 1-\alpha\big) + \delta\big) \\ & \leq \limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,\text{cpt}}}} \big[P_F\big(T_n(\theta) > c_0\big(h_{n,F}(\theta), 1-\alpha\big) + \delta\big) \\ & \quad - P\big(T\big(h_{n,F}(\theta)\big) > c_0\big(h_{n,F}(\theta), 1-\alpha\big)\big)\big] \\ & \quad + \limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,\text{cpt}}}} P\big(T\big(h_{n,F}(\theta)\big) > c_0\big(h_{n,F}(\theta), 1-\alpha\big)\big) \\ & \leq 0 + \alpha,\end{aligned}$$

where the second inequality holds by Theorem 1(a) with $x_{h_{n,F}(\theta)} = c_0(h_{n,F}(\theta), 1-\alpha) + \delta$ and by the definition of the quantile $c_0(h_{n,F}(\theta), 1-\alpha)$ of $T(h_{n,F}(\theta))$. *Q.E.D.*

PROOF OF LEMMA A3: Let $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ be a sequence for which $h_{2,F_n}(\theta_n) \in \mathcal{H}_{2,\text{cpt}}$ and the probability in the statement of the lemma evaluated at (θ_n, F_n) differs from its supremum over $(\theta, F) \in \mathcal{F}$ (with $h_{2,F}(\theta) \in \mathcal{H}_{2,\text{cpt}}$) by δ_n or less, where $0 < \delta_n \to 0$ as $n \to \infty$. Such a sequence always exists. It suffices to show that

$$\begin{aligned}(12.16)\quad & \lim_{n\to\infty} P_{F_n}\big(c\big(\varphi_n(\theta_n), \widehat{h}_{2,n}(\theta_n), 1-\alpha\big) \\ & \quad < c\big(h_{1,n,F_n}(\theta_n), \widehat{h}_{2,n}(\theta_n), 1-\alpha\big)\big) = 0.\end{aligned}$$

By the compactness of $\mathcal{H}_{2,\mathrm{cpt}}$, given any subsequence $\{u_n : n \geq 1\}$ of $\{n\}$, there exists a subsubsequence $\{a_n : n \geq 1\}$ for which $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$ as $n \to \infty$, for some $h_{2,0} \in \mathcal{H}_{2,\mathrm{cpt}}$. This and $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ for all $n \geq 1$ imply that $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in \mathrm{SubSeq}(h_{2,0})$. Hence, it suffices to show that

$$\begin{aligned}(12.17)\quad \lim_{n\to\infty} P_{F_{a_n}}\Big(c\big(\varphi_{a_n}(\theta_{a_n}), \widehat{h}_{2,a_n}(\theta_{a_n}), 1-\alpha\big) \\ < c\big(h_{1,a_n,F_{a_n}}(\theta_{a_n}), \widehat{h}_{2,a_n}(\theta_{a_n}), 1-\alpha\big)\Big) = 0\end{aligned}$$

for $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in \mathrm{SubSeq}(h_{2,0})$.

By Lemma A1(a), for $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in \mathrm{SubSeq}(h_{2,0})$, we have

$$(12.18)\quad \nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot) \Rightarrow \nu_{h_{2,0}}(\cdot) \quad \text{as} \quad n \to \infty.$$

We now show that, for all sequences $\tau_n \to \infty$ as $n \to \infty$, we have

$$(12.19)\quad \lim_{n\to\infty} P_{F_{a_n}}\Big(\sup_{g\in\mathcal{G}, j\leq p} \big|\nu_{a_n,F_{a_n},j}(\theta_{a_n}, g)\big| > \tau_{a_n}\Big) = 0,$$

where $\nu_{a_n,F_{a_n},j}(\theta_{a_n}, g)$ denotes the jth element of $\nu_{a_n,F_{a_n}}(\theta_{a_n}, g)$. We show this by noting that (12.18) and the continuous mapping theorem give: $\forall \tau > 0$,

$$(12.20)\quad \lim_{n\to\infty} P_{F_{a_n}}\Big(\sup_{g\in\mathcal{G}, j\leq p} \big|\nu_{a_n,F_{a_n},j}(\theta_{a_n}, g)\big| > \tau\Big) = P\Big(\sup_{g\in\mathcal{G}, j\leq p} \big|\nu_{h_{2,0},j}(g)\big| > \tau\Big),$$

where $\nu_{h_{2,0},j}(g)$ denotes the jth element of $\nu_{h_{2,0}}(g)$. In addition, the sample paths of $\nu_{h_{2,0},j}(\cdot)$ are bounded a.s., which yields $1(\sup_{g\in\mathcal{G},j\leq p} |\nu_{h_{2,0},j}(g)| > \tau) \to 0$ as $\tau \to \infty$ a.s. Hence, by the bounded convergence theorem,

$$(12.21)\quad \lim_{\tau\to\infty} P\Big(\sup_{g\in\mathcal{G}, j\leq p} \big|\nu_{h_{2,0},j}(g)\big| > \tau\Big) = 0.$$

Equations (12.20) and (12.21) imply (12.19).

Next, we have

$$\begin{aligned}(12.22)\quad \xi_{a_n}(\theta_{a_n}, g) &= \kappa_{a_n}^{-1}\big(\bar{D}_{a_n}^{-1/2}(\theta_{a_n}, g) D_{F_{a_n}}^{1/2}(\theta_{a_n})\big) \\ &\quad \times a_n^{1/2} D_{F_{a_n}}^{-1/2}(\theta_{a_n}) \bar{m}_{a_n}(\theta_{a_n}, g) \\ &= \kappa_{a_n}^{-1} \mathrm{Diag}^{-1/2}\big(\bar{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g)\big) \\ &\quad \times \big(\nu_{a_n,F_{a_n}}(\theta_{a_n}, g) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g)\big),\end{aligned}$$

where the second equality holds by the definitions of $\bar{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g)$, $\nu_{a_n,F_{a_n}}(\theta_{a_n}, g)$, and $h_{1,a_n,F_{a_n}}(\theta_{a_n}, g)$ in (5.2) and $\bar{D}_n(\theta, g) = \mathrm{Diag}(\bar{\Sigma}_n(\theta, g))$.

Consider constants $\{\tau_n : n \geq 1\}$ such that $\tau_n \to \infty$ and $\tau_n/\kappa_n \to 0$ as $n \to \infty$. We have

$$
\begin{aligned}
(12.23)\quad & P_{F_{a_n}}\bigl(c\bigl(\varphi_{a_n}(\theta_{a_n}), \widehat{h}_{2,a_n}(\theta_{a_n}), 1-\alpha\bigr) \\
&\quad < c\bigl(h_{1,a_n,F_{a_n}}(\theta_{a_n}), \widehat{h}_{2,a_n}(\theta_{a_n}), 1-\alpha\bigr)\bigr) \\
&\leq P_{F_{a_n}}\bigl(\varphi_{a_n,j}(\theta_{a_n}, g) > h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g) \\
&\quad \text{for some } j \leq p, \text{ some } g \in \mathcal{G}\bigr) \\
&\leq P_{F_{a_n}}\bigl(\xi_{a_n,j}(\theta_{a_n}, g) > 1 \;\&\; h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g) < B_{a_n} \\
&\quad \text{for some } j \leq p, \text{ some } g \in \mathcal{G}\bigr) \\
&\leq P_{F_{a_n}}\bigl(\bigl[\bar{h}^{-1/2}_{2,a_n,F_{a_n},j}(\theta_{a_n}, g)\nu_{a_n,F_{a_n},j}(\theta_{a_n}, g) \\
&\quad + \bar{h}^{-1/2}_{2,a_n,F_{a_n},j}(\theta_{a_n}, g)h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g)\bigr] > \kappa_{a_n} \\
&\quad \& \; h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g) < B_{a_n} \text{ for some } j \leq p, \text{ some } g \in \mathcal{G}\bigr) \\
&\leq P_{F_{a_n}}\bigl(\bigl[\tau_{a_n} + \bar{h}^{-1/2}_{2,a_n,F_{a_n},j}(\theta_{a_n}, g)h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g)\bigr] > \kappa_{a_n} \;\& \\
&\quad h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g) < B_{a_n} \text{ for some } j \leq p, \text{ some } g \in \mathcal{G}\bigr) \\
&\quad + P_{F_{a_n}}\Bigl(\sup_{g\in\mathcal{G}, j\leq p} \bigl|\bar{h}^{-1/2}_{2,a_n,F_{a_n},j}(\theta_{a_n}, g)\nu_{a_n,F_{a_n},j}(\theta_{a_n}, g)\bigr| > \tau_{a_n}\Bigr) \\
&\leq P_{F_{a_n}}\bigl(\bar{h}^{-1/2}_{2,a_n,F_{a_n},j}(\theta_{a_n}, g)h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g) > \kappa_{a_n} - \tau_{a_n} \;\& \\
&\quad \bar{h}^{-1/2}_{2,a_n,F_{a_n},j}(\theta_{a_n}, g)h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g) < \varepsilon^{-1/2}\bigl(1 + o_p(1)\bigr)B_{a_n} \\
&\quad \text{for some } j \leq p, \text{ some } g \in \mathcal{G}\bigr) + o(1) \\
&= o(1),
\end{aligned}
$$

where the first inequality holds because $c_0(h, 1-\alpha+\eta)$ and $c(h, 1-\alpha)$ are non-increasing in the first p elements of h_1 by Assumption S1(b), the second inequality holds because $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ implies that $h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g) \geq 0$ $\forall j \leq p, \forall g \in \mathcal{G}$ and Assumption GMS1(a) implies that (i) $\varphi_{a_n,j}(\theta_{a_n}, g) = 0 \leq h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g)$ whenever $\xi_{a_n,j}(\theta_{a_n}, g) \leq 1$ and (ii) $\varphi_{a_n,j}(\theta_{a_n}, g) \leq B_{a_n}$ a.s. $\forall j \leq p, \forall g \in \mathcal{G}$, the third inequality holds by (12.22), the fourth inequality holds because $P(A) \leq P(A \cap B) + P(B^c)$, the last inequality holds because (i) $\bar{h}^{-1/2}_{2,a_n,F_{a_n},j}(\theta_{a_n}, g) \leq \varepsilon^{-1/2}h^{-1/2}_{2,0,j}(1_k, 1_k)(1 + o_p(1)) = \varepsilon^{-1/2}(1 + o_p(1))$ by Lemma A1(b) and (5.2) and (ii) the second summand on the left-hand side of the last inequality is $o(1)$ by (12.19) with τ_{a_n} replaced by $\varepsilon^{1/2}\tau_{a_n}/2$ using (i), and the equality holds because $(\kappa_{a_n} - \tau_{a_n}) - \varepsilon^{-1/2}(1 + o_p(1))B_{a_n} = \kappa_{a_n}(1 - \tau_{a_n}/\kappa_{a_n} - \varepsilon^{-1/2}(1 + o_p(1))B_{a_n}/\kappa_{a_n}) = \kappa_{a_n}(1 + o_p(1))$ using Assumption GMS1(b) and $\kappa_{a_n} \to \infty$ as $n \to \infty$.

Hence, (12.17) holds and the lemma is proved. *Q.E.D.*

PROOF OF LEMMA A4: The result of the lemma is equivalent to

$$\begin{aligned}(12.24)\quad &\lim_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,\mathrm{cpt}}}} P_F\big(c_0\big(h_{1,n,F}(\theta),\widehat{h}_{2,n}(\theta),1-\alpha+\eta\big)\\ &\qquad < c_0\big(h_{1,n,F}(\theta),h_{2,F}(\theta),1-\alpha\big)-\varepsilon^*\big)=0,\end{aligned}$$

where $\varepsilon^*=\eta-\delta>0$. By considering a sequence $\{(\theta_n,F_n)\in\mathcal{F}:n\ge 1\}$ that is within $\delta_n\to 0$ of the supremum in (12.24) for all $n\ge 1$, it suffices to show that

$$\begin{aligned}(12.25)\quad &\lim_{n\to\infty} P_{F_n}\big(c_0\big(h_{1,n,F_n}(\theta_n),\widehat{h}_{2,n}(\theta_n),1-\alpha+\eta\big)\\ &\qquad < c_0\big(h_{1,n,F_n}(\theta_n),h_{2,F_n}(\theta_n),1-\alpha\big)-\varepsilon^*\big)=0.\end{aligned}$$

Given any subsequence $\{u_n\}$ of $\{n\}$, there exists a subsubsequence $\{a_n\}$ such that $d(h_{2,F_{a_n}}(\theta_{a_n}),h_{2,0})\to 0$ as $n\to\infty$ for some $h_{2,0}\in\mathcal{H}_{2,\mathrm{cpt}}$ because $h_{2,F_n}(\theta_n)\in\mathcal{H}_{2,\mathrm{cpt}}$. Hence, it suffices to show that (12.25) holds with a_n in place of n.

The condition $d(h_{2,F_{a_n}}(\theta_{a_n}),h_{2,0})\to 0$ and $(\theta_n,F_n)\in\mathcal{F}$ for all $n\ge 1$ imply that $\{(\theta_{a_n},F_{a_n}):n\ge 1\}\in \mathrm{SubSeq}(h_{2,0})$. Hence, by Lemma A1(b), $d(\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}),h_{2,0})\to_p 0$ as $n\to\infty$. Furthermore,

$$\begin{aligned}(12.26)\quad &\widehat{h}_{2,a_n}\big(\theta_{a_n},g,g^*\big)\\ &= \widehat{D}_{a_n}^{-1/2}(\theta_{a_n})\widehat{\Sigma}_{a_n}\big(\theta_{a_n},g,g^*\big)\widehat{D}_{a_n}^{-1/2}(\theta_{a_n})\\ &= \mathrm{Diag}\big(\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n},1_k)\big)^{-1/2}\widehat{h}_{2,a_n,F_{a_n}}\big(\theta_{a_n},g,g^*\big)\\ &\quad\times \mathrm{Diag}\big(\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n},1_k)\big)^{-1/2}.\end{aligned}$$

Hence, $d(\widehat{h}_{2,a_n}(\theta_{a_n}),h_{2,0})\to_p 0$ as $n\to\infty$. Given this, using the almost sure representation theorem as above, we can construct $\{\tilde{h}_{2,a_n}(g,g^*):g,g^*\in\mathcal{G}\}$ such that $d(\tilde{h}_{2,a_n},h_{2,0})\to 0$ as $n\to\infty$ a.s. and $\tilde{h}_{2,a_n}$ and $\widehat{h}_{2,a_n}(\theta_{a_n})$ have the same distribution under (θ_{a_n},F_{a_n}) for all $n\ge 1$.

For fixed ω in the underlying probability space such that $d(\tilde{h}_{2,a_n}(\cdot,\cdot)(\omega),h_{2,0})\to 0$ as $n\to\infty$, Lemma A5 with $h_{2,n}=\tilde{h}_{2,a_n}(\omega)$ $(=\tilde{h}_{2,a_n}(\cdot,\cdot)(\omega))$, $h_{2,n}^*=h_{2,F_{a_n}}(\theta_{a_n})$, $h_{2,0}=h_{2,0}$, and $\eta_1=\eta$ gives: for all $\delta>0$,

$$\begin{aligned}(12.27)\quad &\liminf_{n\to\infty}\big[c_0\big(h_{1,a_n,F_{a_n}}(\theta_{a_n}),\tilde{h}_{2,a_n}(\omega),1-\alpha+\eta\big)+\delta\\ &\qquad - c_0\big(h_{1,a_n,F_{a_n}}(\theta_{a_n}),h_{2,F_{a_n}}(\theta_{a_n}),1-\alpha\big)\big]\ge 0.\end{aligned}$$

Equation (12.27) holds a.s. This implies that (12.25) holds with a_n in place of n because (i) $\tilde{h}_{2,a_n}$ and $\widehat{h}_{2,a_n}(\theta_{a_n})$ have the same distribution for all $n \geq 1$ and (ii) for any sequence of sets $\{A_n : n \geq 1\}$, $P(A_n \text{ ev.})$ $(= P(\bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} A_k)) = 1$ (where ev. abbreviates eventually) implies that $P(A_n) \to 1$ as $n \to \infty$. *Q.E.D.*

PROOF OF LEMMA A5: Below we show that, for $\{h_{2,n}\}$ and $\{h_{2,n}^*\}$ as in the statement of the lemma, for all constants $x_{h_1, h_{2,n}^*} \in R$ that may depend on $h_1 \in \mathcal{H}_1$ and $h_{2,n}^*$, and all $\delta > 0$,

$$\begin{aligned} (12.28)\quad & \limsup_{n\to\infty} \sup_{h_1 \in \mathcal{H}_1} \big[P\big(T(h_1, h_{2,n}) \leq x_{h_1, h_{2,n}^*}\big) \\ & \quad - P\big(T\big(h_1, h_{2,n}^*\big) \leq x_{h_1, h_{2,n}^*} + \delta\big)\big] \leq 0. \end{aligned}$$

Note that this result is similar to the results of Theorem 1.

We use (12.28) to obtain: for all $\delta > 0$ and $\eta_1 > 0$,

$$\begin{aligned} (12.29)\quad & \limsup_{n\to\infty} \sup_{h_1 \in \mathcal{H}_1} P\big(T(h_1, h_{2,n}) \leq c_0\big(h_1, h_{2,n}^*, 1-\alpha\big) - \delta\big) \\ & \leq \limsup_{n\to\infty} \sup_{h_1 \in \mathcal{H}_1} \big[P\big(T(h_1, h_{2,n}) \leq c_0\big(h_1, h_{2,n}^*, 1-\alpha\big) - \delta\big) \\ & \quad - P\big(T\big(h_1, h_{2,n}^*\big) \leq c_0\big(h_1, h_{2,n}^*, 1-\alpha\big) - \delta/2\big)\big] \\ & \quad + \limsup_{n\to\infty} \sup_{h_1 \in \mathcal{H}_1} P\big(T\big(h_1, h_{2,n}^*\big) \leq c_0\big(h_1, h_{2,n}^*, 1-\alpha\big) - \delta/2\big) \\ & \leq 0 + 1 - \alpha \\ & < 1 - \alpha + \eta_1, \end{aligned}$$

where the second inequality holds by (12.28) with $\delta/2$ in place of δ and $x_{h_1, h_{2,n}^*} = c_0(h_1, h_{2,n}^*, 1-\alpha) - \delta$ and by the definition of the $1-\alpha$ quantile of $T(h_1, h_{2,n}^*)$.

We now use (12.29) to show by contradiction that the result of the lemma holds. Suppose the result of the lemma does not hold. Then, there exist constants $\delta > 0$ and $\varepsilon^* > 0$, a subsequence $\{a_n : n \geq 1\}$, and a sequence $\{h_{1,a_n} \in \mathcal{H}_1 : n \geq 1\}$ such that

$$(12.30)\quad \lim_{n\to\infty} \big[c_0(h_{1,a_n}, h_{2,a_n}, 1-\alpha+\eta_1) + \delta - c_0\big(h_{1,a_n}, h_{2,a_n}^*, 1-\alpha\big)\big] \leq -\varepsilon^* < 0.$$

Using this and (12.29), we have

$$\begin{aligned} (12.31)\quad & \limsup_{n\to\infty} P\big(T(h_{1,a_n}, h_{2,a_n}) \leq c_0(h_{1,a_n}, h_{2,a_n}, 1-\alpha+\eta_1) + \delta\big) \\ & \leq \limsup_{n\to\infty} P\big(T(h_{1,a_n}, h_{2,a_n}) \leq c_0\big(h_{1,a_n}, h_{2,a_n}^*, 1-\alpha\big) - \varepsilon^*/2\big) \end{aligned}$$

$$\leq \limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} P\big(T(h_1, h_{2,a_n}) \leq c_0\big(h_1, h^*_{2,a_n}, 1-\alpha\big) - \varepsilon^*/2\big)$$
$$< 1 - \alpha + \eta_1,$$

where the first inequality holds by (12.30) and the last inequality holds by (12.29) with $\varepsilon^*/2$ in place of δ.

Equation (12.31) is a contradiction to (12.30) because the left-hand-side quantity in (12.31) (without the $\limsup_{n\to\infty}$) is greater than or equal to $1-\alpha+\eta_1$ for all $n \geq 1$ by the definition of the $1-\alpha+\eta_1$ quantile $c_0(h_{1,a_n}, h_{2,a_n}, 1-\alpha+\eta_1)$ of $T(h_{1,a_n}, h_{2,a_n})$. This completes the proof of the lemma except for establishing (12.28).

To establish (12.28), we write

$$\begin{aligned}
(12.32)\quad & \limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} \big[P\big(T(h_1, h_{2,n}) \leq x_{h_1,h^*_{2,n}}\big) - P\big(T\big(h_1, h^*_{2,n}\big) \leq x_{h_1,h^*_{2,n}} + \delta\big)\big] \\
&\leq \limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} \big[P\big(T(h_1, h_{2,n}) \leq x_{h_1,h^*_{2,n}}\big) - P\big(T(h_1, h_{2,0}) \\
&\quad \leq x_{h_1,h^*_{2,n}} + \delta/2\big)\big] \\
&\quad + \limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} \big[P\big(T(h_1, h_{2,0}) \leq x_{h_1,h^*_{2,n}} + \delta/2\big) - P\big(T\big(h_1, h^*_{2,n}\big) \\
&\quad \leq x_{h_1,h^*_{2,n}} + \delta\big)\big].
\end{aligned}$$

The first summand on the right-hand side of (12.32) is less than or equal to 0 by the same argument as used to prove Theorem 1(a) with $\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot)$ replaced by $\nu_{h_{2,a_n}}(\cdot)$ in (12.3), where $\nu_{h_{2,a_n}}(\cdot)$ is defined in (4.2), because $d(h_{2,a_n}, h_{2,0}) \to 0$ as $n \to \infty$ implies that the Gaussian processes $\nu_{h_{2,a_n}}(\cdot) \Rightarrow \nu_{h_{2,0}}(\cdot)$ as $n \to \infty$. This argument uses Assumption S2.

Similarly, the second summand on the right-hand side of (12.32) is less than or equal to 0 by an argument analogous to that for Theorem 1(b). Hence, (12.28) is established, which completes the proof. *Q.E.D.*

13. SUPPLEMENTAL APPENDIX B

13.1. *Kolmogorov–Smirnov and Approximate CvM Tests and CS's*

In this appendix, we provide results for Kolmogorov–Smirnov (KS) and approximate CvM (A-CvM) tests and CS's defined in Sections 3.1, 3.5, and 4.2, respectively. A-CvM tests are Cramér–von Mises-type tests in which the test statistic is an infinite sum that is truncated to include only the first s_n functions $\{g_1, \ldots, g_{s_n}\}$, or the test statistic is an integral with respect to the measure Q and the integral is approximated by a (possibly weighted) average over the functions $\{g_1, \ldots, g_{s_n}\}$, which are obtained by simulation or by a quasi-Monte

Carlo method. The same functions $\{g_1, \ldots, g_{s_n}\}$ are used for the test statistic and the critical value. In the case of simulated functions, the probabilistic results given here are for fixed (i.e., nonrandom) functions $\{g_1, \ldots, g_{s_n}\}$. If $\{g_1, \ldots, g_{s_n}\}$ are obtained via i.i.d. draws from Q, then the probability results are made conditional on the observed functions $\{g_1, \ldots, g_{s_n}\}$ for $n \geq 1$.

We show that (i) KS and A-CvM CS's have uniform asymptotic coverage probabilities that are greater than or equal to their nominal level $1-\alpha$, (ii) KS and A-CvM tests have asymptotic power equal to 1 for all fixed alternatives, and (iii) KS and A-CvM tests have asymptotic power that is arbitrarily close to 1 for a broad array of $n^{-1/2}$-local alternatives whose localization parameter is arbitrarily large.

We consider a slightly more general KS statistic than that defined in (3.7):

$$T_n(\theta) = \sup_{g \in \mathcal{G}_n} S\big(n^{1/2}\bar{m}_n(\theta, g), \bar{\Sigma}_n(\theta, g)\big), \tag{13.1}$$

where $\mathcal{G}_n \subset \mathcal{G}$.

For KS tests and CS's, we make use of the following assumptions.

ASSUMPTION KS: *$\mathcal{G}_n \uparrow \mathcal{G}$ as $n \to \infty$.*

Let $\mathcal{W}_{\mathrm{bd}}$ denote a subset of $\mathcal{W}$ (the set of $k \times k$ positive-definite matrices) containing matrices whose eigenvalues are bounded away from zero and infinity.

ASSUMPTION S2′: *$S(m, \Sigma)$ is uniformly continuous in the sense that, for all bounded sets $\mathcal{M}$ in R^k and all sets $\mathcal{W}_{\mathrm{bd}}$,*

$$\sup_{\mu \in [0,\infty)^p \times \{0\}^v} \ \sup_{\substack{m, m_0 \in \mathcal{M}: \\ \|m - m_0\| \leq \delta}} \ \sup_{\substack{\Sigma, \Sigma_0 \in \mathcal{W}_{\mathrm{bd}}: \\ \|\Sigma - \Sigma_0\| \leq \delta}} \big|S(m+\mu, \Sigma) - S(m_0 + \mu, \Sigma_0)\big|$$
$$\to 0 \quad as \quad \delta \to 0.$$

The following lemma shows that Assumption S2′ is not restrictive.

LEMMA B1: *The functions S_1, S_2, and S_3 satisfy Assumption S2′.*

The following assumption is a strengthening of Assumptions LA1(b) and LA2.

ASSUMPTION LA2′: (a) *For all $B < \infty$, $\sup_{g \in \mathcal{G}: h_1(g) \leq B} \|h_{1,n,F_n}(\theta_n, g) - h_1(g)\| \to 0$ as $n \to \infty$, where θ_n, F_n, and $h_1(g)$ are as in Assumption* LA1, *and*

(b) *the* $k \times d$ *matrix* $\Pi_F(\theta, g) = (\partial/\partial\theta')[D_F^{-1/2}(\theta)E_F m(W_i, \theta, g)]$ *exists and satisfies*: *for all sequences* $\{\delta_n : n \geq 1\}$ *such that* $\delta_n \to 0$ *as* $n \to \infty$,

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \sup_{g\in\mathcal{G}} \big\|\Pi_{F_n}(\theta, g) - \Pi_{F_0}(\theta_0, g)\big\| \to 0 \quad as \quad n \to \infty \quad and$$

$$\sup_{g\in\mathcal{G}} \big\|\Pi_{F_0}(\theta_0, g)\big\| < \infty,$$

where θ_0, F_0, *and* F_n *are as in Assumption* LA1.

Assumption LA2′(a) only requires uniform convergence of $h_{1,n,F_n}(\theta_n, g)$ to $h_1(g)$ over $\{g \in \mathcal{G} : h_1(g) \leq B\}$ because uniform convergence over $g \in \mathcal{G}$ typically does not hold. Assumption LA2′ is not restrictive.

For A-CvM tests and CS's, we use Assumptions S2′, LA2′, and the following assumptions, which hold automatically in the case of an approximate test statistic that is a truncated sum with $s_n \to \infty$.

ASSUMPTION A1: *The functions* $\{g_1, \ldots, g_{s_n}\}$ *for* $n \geq 1$ *are fixed* (*i.e.*, *nonrandom*) *and* $s_n \to \infty$ *as* $n \to \infty$.

ASSUMPTION A2: *The functions* $\{g_1, g_2, \ldots\}$ *satisfy*

$$\sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S\big(m^*(g_\ell), h_{2,F_0}(\theta_*, g_\ell) + \varepsilon I_k\big) \to \int S\big(m^*(g), h_{2,F_0}(\theta_*, g) + \varepsilon I_k\big)\, dQ(g) \quad as \quad n \to \infty,$$

where $m^*(g) = (m_1^*(g), \ldots, m_k^*(g))'$, $m_j^*(g) = E_{F_0} m_j(W_i, \theta_*) g_j(X_i)/\sigma_{F_0,j}(\theta_*)$, θ_* *and* F_0 *are defined as in Assumption* FA, $w_{Q,n}(\ell) = Q(\{g_\ell\})$ *in the case of an approximate test statistic that is a truncated sum*, $w_{Q,n}(\ell) = n^{-1}$ *in the case of an approximate test statistic that is a simulated integral*, *and* $w_{Q,n}(\ell)$ *is a suitable weight when a test statistic is approximated by a quasi-Monte Carlo method*.

ASSUMPTION A3: *The functions* $\{g_1, g_2, \ldots\}$ *satisfy*: *for some sequence of constants* $\{B_c^* < \infty : c = 1, 2, \ldots\}$ *such that* $B_c^* \to \infty$ *as* $c \to \infty$,

$$\sum_{\ell=1}^{s_n} w_{Q,n}(\ell) 1\big(h_1(g_\ell) < B_c^*\big) S\big(\Pi_0(g_\ell)\lambda_0, h_2(g_\ell) + \varepsilon I_k\big) \to \int 1\big(h_1(g) < B_c^*\big) S\big(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k\big)\, dQ(g) \quad as \quad n \to \infty,$$

where $\Pi_0(g) = \Pi_{F_0}(\theta_0, g)$, $h_2(g) = h_{2,F_0}(\theta_0, g)$, *and* θ_0 *and* F_0 *are defined as in Assumption* LA1.

Assumptions A1–A3 are not restrictive because (i) they hold automatically if the approximate test statistic is a truncated sum and (ii) if the approximate test statistic is a simulated integral and $\{g_1, g_2, \ldots\}$ are i.i.d. with distribution Q and $s_n \to \infty$ as $n \to \infty$, then they hold conditional on $\{g_1, g_2, \ldots\}$ with probability 1.

The following result establishes that nominal $1-\alpha$ KS and A-CvM CS's have uniform asymptotic coverage probability greater than or equal to $1-\alpha$.

THEOREM B1: *Suppose Assumptions* M, S1, *and* S2′ *hold and Assumption* GMS1 *holds when considering GMS CS's*. *Then*, *for every compact subset* $\mathcal{H}_{2,\mathrm{cpt}}$ *of* $\mathcal{H}_2$, *KS-GMS*, *KS-PA*, *A-CvM-GMS*, *and A-CvM-PA confidence sets* CS_n *satisfy*

$$\liminf_{n\to\infty} \inf_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,\mathrm{cpt}}}} P_F(\theta\in CS_n) \geq 1-\alpha.$$

COMMENTS: (i) None of Assumptions KS, A1, A2, or A3 are needed in Theorem B1.

(ii) Theorem B1 is an analogue of Theorem 2(a) for CS's based on KS and A-CvM statistics. It is proved by making adjustments to the proof of Theorem 2(a). An analogue of Theorem 2(b) is not given here because the proof of Theorem 2(b) does not go through with KS or A-CvM test statistics. The proof of Theorem 2(b) utilizes the bounded convergence theorem, which applies only if the test statistic is an integral with respect to some measure Q. The continuous mapping theorem cannot be applied because the convergence of $h_{1,n,F_n}(\theta_n, g)$ to $h_{1,\infty,F_0}(\theta_0, g)$ is not uniform over $g\in\mathcal{G}$ for many sequences $\{(\theta_n, F_n)\in\mathcal{F}: n\geq 1\}$, where $(\theta_n, F_n)\to(\theta_0, F_0)$.

The next result shows that KS and A-CvM tests have asymptotic power equal to 1 against all fixed alternatives. This implies that any parameter value outside the identified set is included in a KS or A-CvM CS with probability that goes to zero as $n\to\infty$; see the Comment to Theorem 3.

THEOREM B2: *Suppose Assumptions* FA, CI, Q, S1, S3, *and* S4 *hold*, *Assumption* KS *holds when considering the KS test*, *and Assumptions* A1 *and* A2 *hold when considering A-CvM tests*. *Then*, *the KS-GMS and KS-PA tests satisfy the results of Theorem* 3 *concerning power under fixed alternatives*. *In addition*, *A-CvM-GMS and A-CvM-PA tests*, *respectively*, *satisfy*

(a) $\lim_{n\to\infty} P_{F_0}(\bar{T}_{n,s_n}(\theta_*) > c_{s_n}(\varphi_n(\theta_*), \widehat{h}_{2,n}(\theta_*), 1-\alpha)) = 1$ *and*

(b) $\lim_{n\to\infty} P_{F_0}(\bar{T}_{n,s_n}(\theta_*) > c_{s_n}(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha)) = 1$.

The following result is for $n^{-1/2}$-local alternatives.

THEOREM B3: *Suppose Assumptions* M, S1–S4, S2′, LA1, *and* LA2′ *hold*, *Assumptions* KS *and* LA3 *hold when considering the KS test*, *and Assumptions* A1, A3, *and* LA3′ *hold when considering A-CvM tests*. *Let* $\theta_{n,*} = \theta_{n,*}(\beta) = \theta_n + \beta\lambda_0 n^{-1/2}(1+o(1))$ *be as in Assumption* LA1(a) *with* $\lambda = \beta\lambda_0$ *for some* $\beta > 0$ *and* $\lambda_0 \in R^{d_\theta}$. *Then*, *under* $n^{-1/2}$- *local alternatives*, *the A-CvM-GMS and A-CvM-PA tests*, *respectively*, *satisfy*

(a) $\lim_{\beta\to\infty}\lim_{n\to\infty} P_{F_n}(\bar{T}_{n,s_n}(\theta_{n,*}(\beta)) > c_{s_n}(\varphi_n(\theta_{n,*}(\beta)), \widehat{h}_{2,n}(\theta_{n,*}(\beta)), 1-\alpha)) = 1$ *provided Assumption* GMS1 *also holds*,

(b) $\lim_{\beta\to\infty}\lim_{n\to\infty} P_{Fn}(\bar{T}_{n,s_n}(\theta_{n,*}(\beta)) > c_{s_n}(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_{n,*}(\beta)), 1-\alpha)) = 1$, *and*

(c) *KS-GMS and KS-PA tests satisfy parts* (a) *and* (b), *respectively*, *with* $\bar{T}_{n,s_n}(\theta_{n,*}(\beta))$ *replaced by* $T_n(\theta_{n,*}(\beta))$ *and with the subscript* s_n *on* $c_{s_n}(\cdot,\cdot,\cdot)$ *deleted*.

COMMENT: Theorem B3 shows that KS and A-CvM tests have power arbitrarily close to 1 for the same $n^{-1/2}$-local alternatives as Cramér–von Mises tests that are based on integrals with respect to a probability measure Q.

13.2. *Instruments and Weight Functions*

In this section, we provide three additional examples of instruments $\mathcal{G}$ and weight functions Q that satisfy Assumptions CI, M, FA(e), and Q. We also specify non-data-dependent methods for transforming a regressor to lie in [0, 1].

If $x \in R$ is known to lie in an open, closed, or half-open interval denoted by $\lfloor c, d \rfloor$, where $-\infty \le c \le d \le \infty$, then one can transform x into [0, 1] via

$$
\begin{aligned}
(13.2)\qquad t(x) &= \frac{x-c}{d-c} \quad \text{if} \quad c > -\infty \ \&\ d < \infty,\\
t(x) &= \frac{e^x}{1+e^x} \quad \text{if} \quad c = -\infty \ \&\ d = \infty,\\
t(x) &= \frac{e^{x-c}-1}{1+e^{x-c}} \quad \text{if} \quad c > -\infty \ \&\ d = \infty,\\
t(x) &= \frac{2e^{x-d}}{1+e^{x-d}} \quad \text{if} \quad c = -\infty \ \&\ d < \infty.
\end{aligned}
$$

Alternatively, a column vector X_i can be transformed first to have sample mean equal to zero and sample variance matrix equal to I_{d_x} (by left multiplication by the inverse of the lower-triangular Cholesky decomposition of the sample covariance matrix of X_i). Then, it can be transformed to lie in $[0, 1]^{dx}$ by applying the standard normal distribution function $\Phi(\cdot)$ element by element. This method is employed in Section 10.4.

EXAMPLE 3—B-splines: A collection of B-splines provides a set $\mathcal{G}$ that satisfies Assumptions CI and M for those (θ, F) for which $E_F(m_j(W_i,\theta)|X_i =$

x) is a continuous function of x for all $j \leq k$. The regressors are transformed to lie in $[0,1]^{d_x}$. We consider normalized cubic B-splines with equally spaced knots on $[0,1]^{d_x}$. (B-splines of other orders also could be considered.) The class of normalized cubic B-splines is a countable set defined by

$$\begin{aligned}
(13.3)\quad & \mathcal{G}_{\text{B-spline}} = \left\{g(x) : g(x) = B_C(x) \cdot 1_k \text{ for } C \in \mathcal{C}_{\text{B-spline}}\right\}, \quad \text{where} \\
& \mathcal{C}_{\text{B-spline}} = \Bigg\{C^*_{a,r} = \bigtimes_{u=1}^{d_x}\left[\left((a_u-1)/(2r), (a_u+3)/(2r)\right] \cap [0,1]\right] \\
& \qquad \in [0,1]^{d_x} : a = (a_1, \ldots, a_{d_x})', \ a_u \in \{-2, -1, \ldots, 2r\} \\
& \qquad \text{for } u = 1, \ldots, d_x \text{ and } r = r_0, r_0+1, \ldots\Bigg\} \quad \text{and} \\
& B_{C^*_{a,r}}(x) = 1\left(x \in C^*_{a,r}\right) \\
& \quad \times \prod_{u=1}^{d_x} \begin{cases} y_u^3/6, & \text{for } x_u \in \left((a_u-1)/(2r), a_u/(2r)\right], \\ \left(-3y_u^3 + 12y_u^2 - 12y_u + 4\right)/6, & \\ & \text{for } x_u \in \left(a_u/(2r), (a_u+1)/(2r)\right], \\ \left(-3z_u^3 + 12z_u^2 - 12z_u + 4\right)/6, & \\ & \text{for } x_u \in \left((a_u+1)/(2r), (a_u+2)/(2r)\right], \\ z_u^3/6, & \text{for } x_u \in \left((a_u+2)/(2r), (a_u+3)/(2r)\right], \\ 0, & \text{otherwise,} \end{cases} \\
& x = (x_1, \ldots, x_{d_x})', \quad y_u = 2rx_u - (a_u - 1), \quad \text{and} \\
& z_u = 4 - y_u \quad \text{for} \quad u = 1, \ldots, d_x,
\end{aligned}$$

for some positive integer r_0; see Schumaker (2007, p. 136). If $d_x = 1$, a B-spline in $\mathcal{G}_{\text{B-spline}}$ has finite support given by the union of four consecutive subintervals each of length $(2r)^{-1}$. If $d_x \geq 1$, a cubic B-spline in $\mathcal{G}_{\text{B-spline}}$ has support on a d_x-dimensional hypercube in $[0,1]^{d_x}$ with edges of length $4 \cdot (2r)^{-1}$.

Note that a bounded continuous product kernel with bounded support could be used in place of B-splines in Example 3.

Weight Function Q for $\mathcal{G}_{\text{B-spline}}$

There is a one-to-one mapping $\Pi_{\text{B-spline}} : \mathcal{G}_{\text{B-spline}} \to AR^*$, where AR^* is defined as AR is defined in Section 3.4 but with $\{-2, -1, \ldots, 2r\}^{d_x}$ in place of $\{1, \ldots, 2r\}^{d_x}$. We take $Q = \Pi^{-1}_{\text{B-spline}} Q_{\text{AR}^*}$, where Q_{AR^*} is a probability measure on AR^*. For example, the uniform distribution on $a \in \{-2, -1, \ldots, 2r\}^{d_x}$ conditional on r and some discrete mass function $\{w(r) : r = r_0, r_0+1, \ldots\}$ on r

gives the test statistic:

$$T_n(\theta) = \sum_{r=r_0}^{\infty} w(r) \times \sum_{a\in\{-2,-1,\dots,2r\}^{d_x}} (2r+3)^{-d_x} S\big(n^{1/2}\bar{m}_n(\theta, g_{a,r}), \bar{\Sigma}_n(\theta, g_{a,r})\big), \tag{13.4}$$

where $g_{a,r}(x) = B_{C^*_{a,r}}(x) \cdot 1_k$ for $C^*_{a,r} \in \mathcal{C}_{\text{B-spline}}$.

EXAMPLE 4—Data-Dependent Boxes: Next, we consider a class of functions $\mathcal{G}_{\text{box,dd}}$ that is designed to be applied with a data-dependent weight function Q defined below. Because this Q only puts positive weight on center points x that are in the support of X_i, it turns out to be necessary to consider boxes with different left and right edge lengths as measured from the "center" point. (See footnote 52 on p. 20 for an explanation.)

We define

$$\mathcal{G}_{\text{box,dd}} = \{g : g(x) = 1(x \in C) \cdot 1_k \text{ for } C \in \mathcal{C}_{\text{box,dd}}\}, \quad \text{where}$$
$$\mathcal{C}_{\text{box,dd}} = \Bigg\{ C_{x,r_1,r_2} = \bigtimes_{u=1}^{d_x} (x_u - r_{1,u}, x_u + r_{2,u}] : x \in \text{Supp}_{F_{X,0}}(X_i), r_{1,u}, r_{2,u} \in (0, \bar{r})\ \forall u \le d_x \Bigg\} \tag{13.5}$$

for some $\bar{r} \in (0, \infty]$, $x = (x_1, \dots, x_{d_x})'$, $r_1 = (r_{1,1}, \dots, r_{1,d_x})'$, $r_2 = (r_{2,1}, \dots, r_{2,d_x})'$, and $\text{Supp}_{F_{X,0}}(X_i)$ denotes the support of X_i when F_0 is the true distribution.

Data-Dependent Q for $\mathcal{G}_{\text{box,dd}}$

There is a one-to-one mapping $\Pi_{\text{box,dd}} : \mathcal{G}_{\text{box,dd}} \to \{(x, r_1, r_2) \in \text{Supp}_{F_{X,0}}(X_i) \times (0, \bar{r})^{2d_x}\}$. Thus, for any probability measure Q^* on $\{(x, r_1, r_2) \in \text{Supp}_{F_{X,0}}(X_i) \times (0, \bar{r})^{2d_x}\}$, $(\Pi_{\text{box,dd}})^{-1} Q^*$ is a valid probability measure on $\mathcal{G}_{\text{box,dd}}$. In this case, the inverse mapping $(\Pi_{\text{box,dd}})^{-1}$ is $(\Pi_{\text{box,dd}})^{-1}[x, r_1, r_2] = g_{x,r_1,r_2}(\cdot) = 1(\cdot \in C_{x,r_1,r_2}) \cdot 1_k$. Let

$$Q^*_{F_{X,0}} = F_{X,0} \times \text{Unif}\Bigg(\bigg(\bigtimes_{u=1}^{d_x} (0, \sigma_{X,u}\bar{r})\bigg)^2\Bigg), \quad \text{where}$$
$$\sigma^2_{X,u} = \text{Var}_{F_{X,0}}(X_{i,u}) \quad \text{for} \quad u = 1, \dots, d_x, \tag{13.6}$$

and $F_{X,0}$ denotes the true distribution of X_i.[52] The scale factors $\sigma_{X,1}, \ldots, \sigma_{X,d_x}$ are included here to make $Q^*_{F_{X,0}}$ equivariant to location and scale changes in X_i. Of course, $F_{X,0}$ and $\{\sigma^2_{X,u} : u \leq d_x\}$ are unknown, so they need to be replaced by estimators. The distribution $F_{X,0}$ can be estimated by the empirical distribution of X_i based on a subsample of size b_n of $\{X_i : i \leq n\}$, denoted by $\widehat{F}_{X,b_n}(\cdot)$. Here we use the empirical distribution based on a subsample, rather than the whole sample, because the computational costs are large when $b_n = n$ and n is large.[53] The variances $\{\sigma^2_{X,u} : u \leq d_x\}$ can be estimated by the sample variances based on $\{X_i : i \leq n\}$, denoted by $\{\widehat{\sigma}^2_{X,n,u} : u = 1, \ldots, d_x\}$. In this case, the test statistic is

$$
\begin{aligned}
(13.7) \qquad & T_n(\theta) \\
&= \int_{R^{d_x}} \int_{(\times_{u=1}^{d_x}(0,\widehat{\sigma}_{X,n,u}\bar{r}))^2} S\big(n^{1/2}\bar{m}_n(\theta, g_{x,r_1,r_2}), \bar{\Sigma}_n(\theta, g_{x,r_1,r_2})\big) \\
&\quad \times \prod_{u=1}^{d_x} (\widehat{\sigma}_{X,n,u}\bar{r})^{-2}\, dr_1\, dr_2\, d\widehat{F}_{X,m_n}(x) \\
&= b_n^{-1} \sum_{i=1}^{b_n} \int_{(\times_{u=1}^{d_x}(0,\widehat{\sigma}_{X,n,u}\bar{r}))^2} S\big(n^{1/2}\bar{m}_n(\theta, g_{X_i,r_1,r_2}), \\
&\quad \bar{\Sigma}_n(\theta, g_{X_i,r_1,r_2})\big)\, dr_1\, dr_2 \prod_{u=1}^{d_x} (\widehat{\sigma}_{X,n,u}\bar{r})^{-2},
\end{aligned}
$$

where g_{x,r_1,r_2} is as above.

When an approximate test statistic $\bar{T}_{n,s_n}(\theta)$ that is a simulated integral is employed (see (3.16) in Section 3.5), it is defined as in (13.7) but with the integral over (r_1, r_2) replaced by an average over $\ell = 1, \ldots, s_n$, the term $\prod_{u=1}^{d_x}(\widehat{\sigma}_{X,n,u}\bar{r})^{-2}$ deleted, and g_{X_i,r_1,r_2} replaced by $g_{X_i,r_{1,\ell},r_{2,\ell}}$, where $\{(r_{1,\ell}, r_{2,\ell}) : \ell = 1, \ldots, s_n\}$

[52]One might think that a natural data-dependent measure Q is $Q^s = \Pi^{-1}_{\text{box}}(F_{X,0} \times \text{Unif}((0, \bar{r})^{d_x}))$, defined on $\mathcal{G}^s_{\text{box}}$, where $\mathcal{G}^s_{\text{box}}$ is defined as $\mathcal{G}_{\text{box}}$ is defined in (3.13) but with R replaced by $\text{Supp}(X_i)$. However, such a Q does not necessarily have support that contains $\mathcal{G}^s_{\text{box}}$ and, hence, the resulting test may not have power against all fixed alternatives. See the following paragraph for details. It is for this reason that $\mathcal{G}_{\text{box,dd}}$ is defined to contain boxes that are asymmetric about their center points.

The probability distribution Q^s on $\mathcal{G}^s_{\text{box}}$ does not necessarily satisfy Assumption Q. To see why, consider a simple example with $d_x = 1$ and $k = 1$. Suppose X_i takes only four values: 0, 1, 2, 3, each with probability 1/4 and $\bar{r} > 1$. Then, for $g_{1,1}(x) = 1(x \in (0, 2]) \in \mathcal{G}^s_{\text{box}}$, we have $\mathcal{B}(g_{1,1}, \delta) = \{g_{1,1}\}$. This holds because if $\omega > 0$, $g_{1,1+\omega}(0) = 1$ but $g_{1,1}(0) = 0$; if $\omega < 0$, $g_{1,1+\omega}(2) = 0$ but $g_{1,1}(2) = 1$; if $\omega > 0$, $g_{2,1+\omega}(3) = 1$ but $g_{1,1}(3) = 0$; and if $\omega < 0$, $g_{2,1+\omega}(1) = 0$ but $g_{1,1}(1) = 1$. The set $\{g_{1,1}\}$ has zero Q^s measure. So, Q^s does not satisfy Assumption Q.

[53]Also, it is easier to establish the asymptotic validity of this procedure when $b_n/n \to 0$ as $n \to \infty$.

are i.i.d. with a $\mathrm{Unif}(\times_{u=1}^{d_x}(0,\widehat{\sigma}_{X,n,u}\bar{r}))^2$ distribution. Alternatively, in this case, one can take $b_n = s_n$, delete the integral over (r_1, r_2), delete the term $\prod_{u=1}^{d_x}(\widehat{\sigma}_{X,n,u}\bar{r})^{-2}$, and replace g_{X_i,r_1,r_2} by $g_{X_i,r_{1,i},r_{2,i}}$, where $\{(r_{1,i}, r_{2,i}) : i = 1, \ldots, s_n\}$ are as above.

EXAMPLE 5—Continuous/Discrete Regressors: The collections $\mathcal{G}_{\text{c-cube}}$ and $\mathcal{G}_{\text{box}}$ (defined in the main paper) and $\mathcal{G}_{\text{B-spline}}$ and $\mathcal{G}_{\text{box,dd}}$ (defined here) can be used with continuous and/or discrete regressors. However, one can design $\mathcal{G}$ to exploit the known support of discrete regressors. Suppose $X_i = (X_{1,i}', X_{2,i}')'$, where $X_{1,i} \in R^{d_{x,1}}$ is a continuous random vector and $X_{2,i} \in R^{d_{x,2}}$ is a discrete random vector that takes values in a countable set $D = \{x_{2,1}, x_{2,2}, \ldots\}$, where $x_{2,u} \in R^{d_{x,2}}$ for all $u \geq 1$. Define the set $\mathcal{G}_{c/d}$ by

$$\mathcal{G}_{c/d} = \{g : g = g_1 g_2,\ g_1 \in \mathcal{G}_1,\ g_d \in \mathcal{G}_D\}, \tag{13.8}$$

where $x = (x_1', x_2')'$, g_1 is an R^k-valued function of x_1, g_2 is an R-valued function of x_2, $\mathcal{G}_1 = \mathcal{G}_{\text{c-cube}}, \mathcal{G}_{\text{box}}, \mathcal{G}_{\text{B-spline}}$, or $\mathcal{G}_{\text{box,dd}}$, with x and d_x replaced by x_1 and $d_{x,1}$, respectively, and $\mathcal{G}_D = \{g_d : g_d(x_2) = 1_{\{d\}}(x_2)\}$ for $d \in D\}$.

Weight Function Q for $\mathcal{G}_{c/d}$

When $\mathcal{G}$ is of the form $\mathcal{G}_{c/d}$, it is natural to take Q to be of the form $Q_1 \times Q_D$, where Q_1 is a probability measure on $\mathcal{G}_1$, such as any of those considered above with x_1 in place of x, and Q_D is a probability measure on D. If D is a finite set, then one may take Q_D to be uniform. For example, when $\mathcal{G}_1 = \mathcal{G}_{\text{box}}$ and Q_D is uniform, the test statistic is

$$\begin{aligned} T_n(\theta) = \frac{1}{\#D} \sum_{d \in D} \int_{[0,1]^{d_{x,1}}} \int_{(0,\bar{r})^{d_{x,1}}} & S\big(n^{1/2}\bar{m}_n(\theta, g_{x_1,r} g_d), \bar{\Sigma}_n(\theta, g_{x_1,r} g_d)\big) \\ & \times \bar{r}^{-d_x}\, dr\, dx_1, \end{aligned} \tag{13.9}$$

where $\#D$ denotes the number of elements in D and $x_1 \in R^{d_{x,1}}$. When $\mathcal{G}_1 = \mathcal{G}_{\text{c-cube}}$ or $\mathcal{G}_{\text{B-spline}}$, $T_n(\theta)$ is a combination of the formulae given above.

The following result establishes Assumptions CI, M, and FA(e) for $\mathcal{G}_{\text{B-spline}}$, $\mathcal{G}_{\text{box,dd}}$, and $\mathcal{G}_{c/d}$ and Assumption Q for the weight functions Q on these sets.

LEMMA B2: (a) *For any moment function $m(W_i, \theta)$, Assumptions* CI *and* M *hold with $\mathcal{G} = \mathcal{G}_{\text{B-spline}}$ for all (θ, F) for which $E_F(m_j(W_i, \theta)|X_i = x)$ is a continuous function of x for all $j \leq k$.*

(b) *For any moment function $m(W_i, \theta)$, Assumptions* CI *and* M *hold with $\mathcal{G} = \mathcal{G}_{\text{box,dd}}$.*

(c) *For any moment function $m(W_i, \theta)$, Assumptions* CI *and* M *hold with $\mathcal{G} = \mathcal{G}_{c/d}$, where $\mathcal{G}_1 = \mathcal{G}_{\text{c-cube}}, \mathcal{G}_{\text{box}}, \mathcal{G}_{\text{B-spline}}$, or $\mathcal{G}_{\text{box,dd}}$, with (x, d_x) replaced by $(x_1, d_{x,1})$ and in the case of $\mathcal{G}_1 = \mathcal{G}_{\text{B-spline}}$ Assumption* CI *and* M *only hold for (θ, F) for* *which* $E_F(m_j(W_i,\theta)|X_{i,1}=x_1, X_{2,i}=d)$ *is a continuous function of* $x_1 \in [0,1]^{d_{x,1}}$ $\forall d \in D, \forall j \le k$.

(d) *Assumption* FA(e) *holds for* $\mathcal{G}_{\text{B-spline}}$, $\mathcal{G}_{\text{box,dd}}$, *and* $\mathcal{G}_{c/d}$.

(e) *Assumption* Q *holds for the weight function* $Q_c = \Pi^{-1}_{\text{B-spline}} Q_{\text{AR}^*}$ *on* $\mathcal{G}_{\text{B-spline}}$, *where* Q_{AR^*} *is uniform on* $a \in \{-2,-1,\dots,2r\}^{d_x}$ *conditional on* r *and* r *has some probability mass function* $\{w(r): r = r_0, r_0+1, \dots\}$ *with* $w(r) > 0$ *for all* r.

(f) *Assumption* Q *holds for the weight function* $Q_d = (\Pi_{\text{box,dd}})^{-1} Q^*_{F_{X,0}}$, *where* $Q^*_{F_{X,0}} = (F_{X,0} \times \text{Unif}((\times_{u=1}^{d_x}(0,\sigma_{X,u}\bar{r}))^2)$ *on* $\mathcal{G}_{\text{box,dd}}$.

(g) *Assumption* Q *holds for the weight function* $Q_e = Q_1 \times Q_D$ *on* $\mathcal{G}_{c/d}$, *where* Q_1 *is a probability measure on* $\mathcal{G}_1$ *equal to any of the distributions* Q *on* $\mathcal{G}$ *considered in part* (e), *part* (f), *or in Lemma* 4 *but with* x_1 *in place of* x, D *is a finite set, and* $Q_D = \text{Unif}(D)$.

COMMENT: The uniform distribution that appears in parts (e)–(g) of the lemma could be replaced by another distribution and the results of the lemma still hold provided the other distribution has the same support. For example, in part (g), Assumption Q holds when D is a countably infinite set and Q_D is a probability measure whose support is D.

13.3. *Sufficient Conditions for Assumption GMS2(a)*

The following lemma verifies Assumption GMS2(a) for the CvM statistic under some conditions on S, Q, and α.

LEMMA B3: *Suppose Assumptions* S1, S3, Q, *and* EP *hold and* S *is the Sum or Max function. Consider any* $(\theta_c, F_c) \in \mathcal{F}$. *Then,*

(a) *the d.f. of* $T(h_{\infty,F_c}(\theta_c))$ *is continuous and strictly increasing at all* $x > 0$ *and*

(b) *if, in addition,* $Q(\mathcal{G}_{h0}) > 0$, *where* $\mathcal{G}_{h0} = \{g \in \mathcal{G} : h_{1,\infty,F_c,j^*}(\theta_c, g) = 0\}$ *for some* $j^* \le k$ *and* $h_{2,F_c,j^*}(g^*) > 0$ *for some* $g^* \in \mathcal{G}_{h0}$, *then the* $1-\alpha$ *quantile of* $T(h_{\infty,F_c}(\theta_c))$ *is positive for any* $\alpha < 1/2$.

COMMENTS: (i) In the case of i.i.d. observations and no preliminary estimator $\widehat{\tau}_n(\theta)$, Assumption EP is implied by Assumption M, so Assumption EP holds under the conditions of Theorem 2(b).

(ii) The proof of Lemma B3 does not go through when S is the QLR function because, in that case, $T(h_{\infty,F_c}(\theta_c))$ is not a convex function of the Gaussian process $\nu_{h_{2,F}}(\cdot)$.

PROOF OF LEMMA B3: When S is the Sum or Max function, $T(h_{\infty,F_c}(\theta_c))$ is a convex function of the Gaussian process $\nu_{h_{2,F}}(\cdot)$. By Theorem 11.1 of Davydov, Lifshits, and Smorodina (1995), the d.f. of $T(h_{\infty,F_c}(\theta_c))$ is continuous and strictly increasing at every point in its support except the left endpoint. To prove part (a), it remains to show that the left endpoint of the support of $T(h_{\infty,F_c}(\theta_c))$ is zero.

It suffices to show that zero is in the support of $T(h_{\infty,F_c}(\theta_c))$ because $T(h_{\infty,F_c}(\theta_c)) \geq 0$ with probability 1 by Assumption S1(c). For any $\zeta > 0$,

$$
\begin{aligned}
(13.10)\quad & \Pr\bigl(T\bigl(h_{\infty,F_c}(\theta_c)\bigr) < \zeta\bigr) \\
& \geq \Pr\Bigl(\sup_{g\in\mathcal{G}} S\bigl(\nu_{h_{2,F_c}}(g), h_{2,F_c}(g) + \varepsilon I_k\bigr) < \zeta\Bigr) \\
& \geq \Pr\Bigl(\sup_{g\in\mathcal{G}, j\leq k} \bigl|\bigl(h_{2,F_c,j}(g) + \varepsilon\bigr)^{-1/2} \nu_{h_{2,F_c},j}(g)\bigr| < \sqrt{\zeta/k}\Bigr) \\
& > 0,
\end{aligned}
$$

where the second inequality holds for both the Sum and Max functions (where, for the Max function, there is no need to divide ζ by k) and the third inequality holds because, by Problem 11.3 of Davydov, Lifshits, and Smorodina (1995, p. 79), zero is the infimum of the support of the supremum of the absolute value of a Gaussian process whose support is the set of bounded continuous functions. Thus, part (a) holds.

Next, we prove part (b). Suppose that $Q(\mathcal{G}_{h0}) > 0$. Then,

$$
\begin{aligned}
(13.11)\quad & \Pr\bigl(T\bigl(h_{\infty,F_c}(\theta_c)\bigr) > 0\bigr) \\
& \geq \Pr\Bigl(\int_{\mathcal{G}_{h0}} S\bigl(\nu_{h_{2,F_c}}(g) + h_{1,\infty,F_c}(\theta_c, g), h_{2,F_c}(g) + \varepsilon I_k\bigr)\, dQ(g) > 0\Bigr) \\
& \geq \Pr\bigl(S\bigl(\nu_{h_{2,F_c}}\bigl(g^*\bigr) + h_{1,\infty,F_c}(\theta_c, g^*), h_{2,F_c}\bigl(g^*\bigr) + \varepsilon I_k\bigr) > 0\bigr) \\
& \geq \Pr\bigl(\nu_{h_{2,F_c},j^*}\bigl(g^*\bigr) < 0\bigr) \\
& = 1/2,
\end{aligned}
$$

where the first inequality holds by Assumption S1(c), the second inequality holds by Assumption Q, $Q(\mathcal{G}_{h0}) > 0$, and the continuity of $\nu_{h_{2,F_c}}(\cdot)$, $h_{2,F_c}(\cdot)$, and S (by Assumption S1(d)), the third inequality holds by Assumption S3 using $h_{1,\infty,F_c,j^*}(\theta_c, g^*) = 0$ (whether or not the j^*th moment condition is an equality or an inequality), and the equality holds because $\nu_{h_{2,F_c},1}(g^*)$ is a normal random variable with mean zero and positive variance $h_{2,F_c,j}(g^*)$. *Q.E.D.*

13.4. *Example: Verification of Assumptions LA1–LA3 and LA3′*

Here we verify Assumptions LA1–LA3 and LA3′ in a simple example for purposes of illustration. These assumptions are the main assumptions employed with local alternatives.

EXAMPLE: Suppose $W_i = (Y_i, X_i)' \in R^2$ and there is a single moment inequality function $m(W_i, \theta) = Y_i - \theta$ and no moment equalities, that is, $p = 1$ and $v = 0$. Suppose the true parameters/distributions $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ and the null values $\{\theta_{n,*} \in \Theta : n \geq 1\}$ satisfy: (i) $\theta_n \to \theta_0$ and $F_n \to F_0$ (under the Kolmogorov metric) for some $(\theta_0, F_0) \in \mathcal{F}$, (ii) $\theta_{n,*} = \theta_n + \lambda n^{-1/2}$ for some $\lambda > 0$, (iii) $Y_i = \theta_n + \mu(X_i)n^{-1/2} + U_i$, (iv) $\mu(x) \geq 0$, $\forall x \in R$, and (v) under all F such that $(\theta, F) \in \mathcal{F}$ for some $\theta \in \Theta$, (X_i, U_i) are i.i.d. with distribution that does not depend on F, X_i and U_i are independent, $E_F U_i = 0$, $\mathrm{Var}_F(U_i) = 1$, $\mathrm{Var}_F(X_i) \in (0, \infty)$, and $E_F|U_i|^{2+\delta} + E_F|\mu(X_i)|^{2+\delta} < \infty$ for some $\delta > 0$, and $\sup_{g \in \mathcal{G}} E_F(1 + \mu^2(X_i))(1 + g^2(X_i)) < \infty$.

We show that, in this example, Assumptions LA1 and LA2 hold, Assumption LA3 holds if λ is sufficiently large, and Assumption LA3′ holds if $\mathcal{G}$ and Q satisfy Assumptions CI and Q, respectively.

By (v), we can write $E_F g(X_i) = Eg(X_i)$ and $E_F \mu(X_i)g(X_i) = E\mu(X_i)g(X_i)$.

Assumption LA1(a) holds by (i) and (ii). Assumption LA1(b) holds by the following calculations:

$$\begin{aligned}
n^{1/2} E_{F_n} m(W_i, \theta_n, g) &= n^{1/2} E_{F_n}\big(U_i + \mu(X_i)n^{-1/2}\big)g(X_i) \\
&= h_1(g), \quad \text{where} \\
h_1(g) &= E\mu(X_i)g(X_i) \in [0, \infty) \quad \text{and} \\
\sigma^2_{F_n}(\theta_n) &= \mathrm{Var}_{F_n}(Y_i) = \mathrm{Var}_{F_n}\big(U_i + \mu(X_i)n^{-1/2}\big) \\
&= 1 + n^{-1}\mathrm{Var}_{F_n}\big(\mu(X_i)\big) \to 1.
\end{aligned} \tag{13.12}$$

To show Assumption LA1(c), we have

$$\begin{aligned}
E_{F_n} Y_i^2 g(X_i)g^*(X_i) &= E_{F_n}\big(\theta_n + \mu(X_i)n^{-1/2} + U_i\big)^2 g(X_i)g^*(X_i) \\
&\to E_{F_0}(\theta_0 + U_i)^2 g(X_i)g^*(X_i) \\
&= E_{F_0} Y_i^2 g(X_i)g^*(X_i) \quad \text{as} \quad n \to \infty,
\end{aligned} \tag{13.13}$$

uniformly over $g, g^* \in \mathcal{G}$, using (i), (iii), and (v). Here we have used $Y_i = \theta_0 + U_i$ under F_0. This holds because $F_n \to F_0$ by (ii), which implies that $P_{F_n}(Y_i \leq y) \to P_{F_0}(Y_i \leq y)$ for all continuity points Y_i, but direct calculations show that $P_{F_n}(Y_i \leq y) = P(\theta_n + \mu(X_i)n^{-1/2} + U_i \leq y) \to P(\theta_0 + U_i \leq y)$ for all continuity points y of $U_i + \theta_0$ and, hence, $Y_i = \theta_0 + U_i$ under F_0.

Next, we write

$$\begin{aligned}
&E_{F_n} m(W_i, \theta_n, g) m\big(W_i, \theta_n, g^*\big) \\
&\quad = E_{F_n} Y_i^2 g(X_i)g^*(X_i) - \theta_n E\big[E_{F_n}(Y_i|X_i)\big(g(X_i) + g^*(X_i)\big)\big] \\
&\qquad + \theta_n^2 E g(X_i)g^*(X_i) \\
&\quad = E_{F_n} Y_i^2 g(X_i)g^*(X_i) - \theta_n E\big[\big(\theta_n + \mu(X_i)n^{-1/2}\big)\big(g(X_i) + g^*(X_i)\big)\big]
\end{aligned} \tag{13.14}$$

$$
\begin{aligned}
&+ \theta_n^2 Eg(X_i)g^*(X_i) \\
&= E_{F_0} Y_i^2 g(X_i)g^*(X_i) - \theta_0^2 Eg(X_i) - \theta_0^2 Eg^*(X_i) \\
&\quad + \theta_0^2 Eg(X_i)g^*(X_i) + o(1) \\
&= E_{F_0} m(W_i, \theta_0, g) m\big(W_i, \theta_0, g^*\big) + o(1),
\end{aligned}
$$

where $o(1)$ holds uniformly over $g, g^* \in \mathcal{G}$, using (13.13), (i), (iii), and (v). In addition, $E_{F_n} m(W_i, \theta_n, g) = o(1)$ and $E_{F_0} m(W_i, \theta_0, g) = o(1)$ uniformly over $g \in \mathcal{G}$ by (13.12) and (v). Hence, the first part of Assumption LA1(c) holds. The second part of Assumption LA1(c) holds by the same argument with $\theta_{n,*}$ in place of θ_n.

Assumption LA1(d) holds because $\mathrm{Var}_{F_n}(m_j(W_i, \theta_{n,*})) = \mathrm{Var}_{F_n}(m_j(W_i, \theta_n)) > 0$. Assumption LA1(e) holds using (v) and the above expression for $\sigma_{F_n}^2(\theta_n)$.

Assumption LA2 holds because $\Pi_F(\theta, g)$ does not depend on (θ, F) by the following calculations and (v): $\forall F$ such that $(\theta, F) \in \mathcal{F}$ and $\forall g \in \mathcal{G}$,

$$
\begin{aligned}
(13.15) \quad \Pi_F(\theta, g) &= (\partial/\partial\theta)\big[D_F^{-1/2}(\theta) E_F m(W_i, \theta, g)\big] \\
&= \sigma_F^{-1}(\theta)(\partial/\partial\theta) E_F (Y_i - \theta) g(X_i) = -\sigma_F^{-1}(\theta) Eg(X_i),
\end{aligned}
$$

where the second equality holds because $D_F(\theta) = \sigma_F^2(\theta) = \mathrm{Var}_F(Y_i)$ does not depend on θ.

We have $\Pi_0(g) = \Pi_{F_0}(\theta_0, g) = -Eg(X_i)$ by (13.15) and $\sigma_{F_0}^2(\theta_0) = 1$. Hence, in Assumption LA3, $h_1(g) + \Pi_0(g)\lambda = E\mu(X_i)g(X_i) - Eg(X_i)\lambda$, which is negative whenever $\lambda > E\mu(X_i)g(X_i)/Eg(X_i)$. Hence, if the null value $\theta_{n,*}$ deviates from the true value θ_n by enough (i.e., if $n^{1/2}(\theta_{n,*} - \theta_n) = \lambda$ is large enough), then the null hypothesis is violated for all n and Assumption LA3 holds.

Next, we show that Assumption LA3′ holds provided Assumptions CI and Q hold. We have: (a) $\Pi_0(g) = -Eg(X_i)$, (b) $h_1(g) < \infty$ $\forall g \in \mathcal{G}$ by (13.12) using (v), and (c) $\lambda_0 = \lambda/\beta > 0$ because $\lambda > 0$ by (ii) and $\beta > 0$ by definition. Hence, the condition of Assumption LA3′ reduces to

$$
(13.16) \quad Q\big(\{g \in \mathcal{G} : Eg(X_i) > 0\}\big) > 0.
$$

Suppose $Eg^*(X_i) > 0$ for some $g^* \in \mathcal{G}$. (This is a very weak requirement on $\mathcal{G}$ and is implied by Assumption CI; see below.) Let $\delta_1 = Eg^*(X_i) > 0$. Then, using the metric ρ_X defined in (6.3), for any $g \in \mathcal{G}$ with $\rho_X(g, g^*) < \delta_1$, we have $Eg(X_i) > 0$ because otherwise $g(X_i) = 0$ a.s. and $\delta_1 > \rho_X(g, g^*) = (Eg^*(X_i)^2)^{1/2} \geq Eg^*(X_i) = \delta_1$, which is a contradiction. Thus, $Eg(X_i) > 0$ for all $g \in \mathcal{B}_{\rho_X}(g^*, \delta_1)$, where $\mathcal{B}_{\rho_X}(g^*, \delta_1)$ is the open ρ_X-ball in $\mathcal{G}$ centered at g^* with radius δ_1. By Assumption Q, $Q(\mathcal{B}_{\rho_X}(g^*, \delta_1)) > 0$. Hence, (13.16) holds and Assumption LA3′ is verified.

Lastly, we show that Assumption CI implies that $Eg^*(X_i) > 0$ for some $g^* \in \mathcal{G}$. For all $\theta > \theta_0$, we have

$$\begin{aligned} (13.17) \quad \mathcal{X}_{F_0}(\theta) &= \{x \in R : E_{F_0}(m_j(W_i, \theta)|X_i = x) < 0\} \\ &= \{x \in R : \theta_0 - \theta < 0\} = R, \end{aligned}$$

where the second equality holds because $Y_i = \theta_0 + U_i$ under F_0, and so, $E_{F_0}(m_j(W_i, \theta)|X_i = x) = E_{F_0}(Y_i - \theta|X_i = x) = \theta_0 - \theta$.

By (13.17), $P_{F_0}(X_i \in \mathcal{X}_{F_0}(\theta)) = P_{F_0}(X_i \in R) = 1 > 0$. Hence, by Assumption CI, there exists $g^* \in \mathcal{G}$ such that $E_{F_0} m(W_i, \theta) g^*(X_i) = E(\theta_0 - \theta) g^*(X_i) < 0$ for $\theta > \theta_0$. That is, $Eg^*(X_i) > 0$.

13.5. *Uniformity Issues With Infinite-Dimensional Nuisance Parameters*

This section illustrates one of the subtleties that arises when considering the uniform asymptotic behavior of a test or CS in a scenario in which a test statistic exhibits a "discontinuity in its asymptotic distribution" and an infinite-dimensional nuisance parameter affects the asymptotic behavior of the test statistic.

In many testing problems, the asymptotic distribution of a KS-type statistic is determined by establishing the weak convergence of some underlying stochastic process and applying the continuous mapping theorem. This yields the asymptotic distribution to be the supremum of the limit process. In the context of conditional moment inequalities with drifting sequences of distributions, this method does not work. The reason is that the normalized mean function of the underlying stochastic process, that is, $h_{1,n,F_n}(\theta_n, g)$, often (in fact, usually) does not converge uniformly over $g \in \mathcal{G}$ to its pointwise limit, that is, $h_1(g)$, and, hence, stochastic equicontinuity fails.[54]

We show by counterexample that the asymptotic distribution under drifting sequences of null distributions of a KS statistic, where the "sup" is over $g \in \mathcal{G}$, does not necessarily equal the supremum of the limiting process indexed by $g \in \mathcal{G}$ that is determined by the finite-dimensional distributions. Hence, if the critical value is based on this limiting process, a KS test does not necessarily have correct asymptotic null rejection probability. In fact, we show that it can over-reject the null hypothesis substantially.

The same phenomenon does not arise with CvM statistics, which are "average" statistics. This is because the averaging smooths out the nonuniform convergence of the normalized mean function.

The results in the first section of this appendix show that the problem discussed above does not arise with the KS statistic when the critical value employed is a GMS critical value that satisfies Assumption GMS1 (see Section 4)

[54]Note that drifting sequences of distributions are of interest because correct asymptotic coverage probabilities under all drifting sequences is necessary, though not sufficient, for correct uniform asymptotic coverage probabilities.

or a PA critical value. The validity of these critical values is established using a uniform asymptotic approximation of the distribution of the KS statistic, rather than using asymptotics under sequences of true distributions.

To start, we give a very simple deterministic example to illustrate a situation in which a deterministic KS statistic does not converge to the supremum of the pointwise limit, but an "average" CvM statistic does converge to the average of the pointwise limit. Consider the piecewise linear functions $f_n : [0, 1] \to [0, 1]$ defined by

$$f_n(x) = \begin{cases} x/\varepsilon_n, & \text{for } x \in [0, \varepsilon_n], \\ 1 - (x - \varepsilon_n)/\varepsilon_n, & \text{for } x \in [\varepsilon_n, 2\varepsilon_n], \\ 0, & \text{for } x \in [2\varepsilon_n, 1], \end{cases} \tag{13.18}$$

where $0 < \varepsilon_n \to 0$ as $n \to \infty$. Then, for all $x \in [0, 1]$,

$$f_n(x) \to f(x) = 0 \quad \text{as} \quad n \to \infty. \tag{13.19}$$

The KS statistic does not converge to the supremum of the limit function:

$$\sup_{x\in[0,1]} f_n(x) = 1 \nrightarrow 0 = \sup_{x\in[0,1]} f(x) \quad \text{as} \quad n \to \infty. \tag{13.20}$$

On the other hand, the CvM statistic does converge to the average of the limit function:

$$\int_0^1 f_n(x)\,dx = \varepsilon_n \to 0 = \int_0^1 f(x)\,dx \quad \text{as} \quad n \to \infty. \tag{13.21}$$

The convergence result for the KS statistic in (13.20) is potentially problematic because, in a testing problem with a KS statistic, the critical value might be obtained from the distribution of the supremum of the limit process. If convergence in distribution of the KS statistic to the "sup" of the limit process does not hold, then such a critical value is not necessarily appropriate.

Now we show that the phenomenon illustrated in (13.18)–(13.21) arises in conditional moment inequality models. We consider a particular conditional moment inequality model with a single linear moment inequality, a fixed true value θ_0, and a particular drifting sequence of distributions. (Note that CX stands for "counterexample.")

ASSUMPTION CX: (a) *$m(W_i, \theta) = Y_i - \theta$ for $Y_i, \theta \in R$,*

(b) *$m(W_i, \theta_0) = Y_i = U_i + 1(X_i \in (\varepsilon_n, 1])$, where the true value θ_0 equals* 0, *$EU_i = 0$, $EU_i^2 = 1$, the distribution of U_i does not depend on n, U_i and X_i are independent, and the constants $\{\varepsilon_n : n \geq 1\}$ satisfy $\varepsilon_n \to 0$ as $n \to \infty$,*

(c) *$X_i = \varepsilon_n$ with probability* $1/2$ *and X_i is uniform on* $[0, 1]$ *with probability* $1/2$,

(d) $\{W_i = (Y_i, X_i)' : i \leq n, n \geq 1\}$ *is a row-wise independent and identically distributed triangular array* (*with the dependence of* W_i, Y_i, *and* X_i *on* n *suppressed for notational simplicity*),
(e) $S(m, \Sigma) = S(m)$ *for* $m \in R$,
(f) S *satisfies Assumptions* S1 *and* S2, *and*
(g) $\mathcal{G} = \{g_{a,b} : g_{a,b} = 1(x \in (a, b])$ *for some* $0 \leq a < b \leq 1\}$.

The function $S_1(m) = [m]_-^2$ satisfies Assumptions CX(e)–(f). Assumption CX(e) is made for simplicity. It could be removed and, with some changes to the proofs, the results given below would hold for $S = S_2$ as well. The class of functions $\mathcal{G}$ specified in Assumption CX(g) is the class of one-dimensional boxes, as in Example 1 of Section 3.3.

We write

$$(13.22) \quad n^{1/2}\bar{m}_n(\theta_0, g_{a,b}) = n^{-1/2}\sum_{i=1}^n Y_i g_{a,b}(X_i) = \nu_n(g_{a,b}) + h_{1,n}(g_{a,b}), \quad \text{where}$$

$$\nu_n(g_{a,b}) = n^{1/2}\big(\bar{m}_n(\theta_0, g_{a,b}) - E_{F_n}\bar{m}_n(\theta_0, g_{a,b})\big) \quad \text{and}$$

$$h_{1,n}(g_{a,b}) = n^{1/2}E_{F_n}\bar{m}_n(\theta_0, g_{a,b}).$$

The KS statistic is

$$(13.23) \quad \sup_{g_{a,b}\in\mathcal{G}} S\big(n^{1/2}\bar{m}_n(\theta_0, g_{a,b})\big) = \sup_{g_{a,b}\in\mathcal{G}} S\big(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})\big).$$[55]

Let $\nu(\cdot)$ be a mean zero Gaussian process indexed by $g_{a,b} \in \mathcal{G}$ with covariance kernel $K(\cdot,\cdot)$ and with sample paths that are uniformly ρ-continuous, where $K(\cdot,\cdot)$ and $\rho(\cdot,\cdot)$ are specified in the proof of Theorem B4 given in the next subsection.

The KS statistic satisfies the following result.

THEOREM B4: *Suppose Assumption* CX *holds. Then*,
(a) $\nu_n(\cdot) \Rightarrow \nu(\cdot)$ *as* $n \to \infty$,
(b) $h_{1,n}(g_{a,b}) \to h_1(g_{a,b}) = \infty$ *as* $n \to \infty$ *for all* $g_{a,b} \in \mathcal{G}$,
(c) $\sup_{g_{a,b}\in\mathcal{G}} |h_{1,n}(g_{a,b}) - h_1(g_{a,b})| \nrightarrow 0$ *as* $n \to \infty$,
(d) $S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})) \to_d S(\nu(g_{a,b}) + h_1(g_{a,b}))$ *as* $n \to \infty$ *for all* $g_{a,b} \in \mathcal{G}$,
(e) $\sup_{g_{a,b}\in\mathcal{G}} S(\nu(g_{a,b}) + h_1(g_{a,b})) = 0$ *a.s.*,
(f) $\sup_{g_{a,b}\in\mathcal{G}} S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})) \geq S(\nu_n(g_{0,\varepsilon_n}) + h_{1,n}(g_{0,\varepsilon_n})) \to_d S(Z^*)$ *as* $n \to \infty$, *where* $Z^* \sim N(0, 1/2)$ *and the inequality holds a.s.*, *and*
(g) $\sup_{g_{a,b}\in\mathcal{G}} S(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})) \nrightarrow_d \sup_{g_{a,b}\in\mathcal{G}} S(\nu(g_{a,b}) + h_1(g_{a,b}))$ *as* $n \to \infty$.

[55]Note that for simplicity we have not rescaled the moment functions $\bar{m}_n(\theta_n, g_{a,b})$ that appear in the definition of the "sup" standard deviation estimators.

COMMENTS: (i) Theorem B4(g) shows that the KS statistic does not have an asymptotic distribution that equals the supremum over $g_{a,b} \in \mathcal{G}$ of the pointwise limit given in Theorem B4(d). This is due to the lack of uniform convergence of $h_{1,n}(g_{a,b})$ shown in Theorem B4(c). (Note that the convergence in part (d) of the theorem also holds jointly over any finite set of $g_{a,b} \in \mathcal{G}$.)

(ii) Let $c_{\infty,1-\alpha}$ denote the $1-\alpha$ quantile of $\sup_{g_{a,b}\in\mathcal{G}} S(\nu(g_{a,b}) + h_1(g_{a,b}))$. By Theorem B4(e), $c_{\infty,1-\alpha} = 0$. Theorem B4(f) and some calculations (given in the proof of Theorem B4 below) yield

$$\liminf_{n\to\infty} P\Big(\sup_{g_{a,b}\in\mathcal{G}} S\big(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})\big) > c_{\infty,1-\alpha}\Big) \geq 1/2. \tag{13.24}$$

That is, if one uses $c_{\infty,1-\alpha}$ as the critical value, the nominal level α test based on the KS statistic has an asymptotic null rejection probability that is bounded below by $1/2$, which indicates substantial over-rejection.

Next, we provide results for a CvM statistic defined by

$$\int S\big(n^{1/2}\bar{m}_n(\theta_0, g_{a,b})\big)\, dQ(g_{a,b}) = \int S\big(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})\big)\, dQ(g_{a,b}), \tag{13.25}$$

where Q is a probability measure on $\mathcal{G}$. In contrast to the KS statistic, the CvM statistic is well-behaved asymptotically.

THEOREM B5: *Suppose Assumption* CX *holds*. *Then*,

$$\int S\big(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})\big)\, dQ(g_{a,b}) \to_d \int S\big(\nu(g_{a,b}) + h_1(g_{a,b})\big)\, dQ(g_{a,b}) \quad \textit{as} \quad n \to \infty.$$

COMMENT: Theorem B5 is not proved using the continuous mapping theorem due to the nonuniform convergence of $h_{1,n}(g_{a,b})$. Rather, it is proved using an almost sure representation argument coupled with the bounded convergence theorem.

13.6. *Problems With Pointwise Asymptotics*

In the case of unconditional moment inequalities, pointwise asymptotics have been shown in Andrews and Guggenberger (2009) to be deficient in the sense that they fail to capture the finite-sample properties of a typical test statistic of interest. This is due to the discontinuity in the asymptotic distribution of the test statistic. In the case of *conditional* moment inequalities, the deficiency of pointwise asymptotics is even greater. We show in a simple example that the asymptotic distribution of a test statistic $T_n(\theta_0)$ under a fixed distribution F_0 often is *pointmass at zero* even when the true parameter θ_0 is on the boundary of the identified set. This does not reflect the statistic's finite-sample distribution.

Suppose (i) $W_i = (Y_i, X_i)'$, (ii) there is one moment inequality function $m(W_i, \theta) = Y_i - \theta$ and no moment equalities (i.e., $p = 1$ and $v = 0$), (iii) the true distribution is F_0 for all $n \geq 1$, (iv) $Y_i = \theta_0 + \mu(X_i) + U_i$, where $X_i, U_i \in R$ and $\mu(\cdot) = \mu_{F_0}(\cdot)$, (v) $\mu(x) \geq 0\ \forall x \in R$, $\mathcal{X}_{\text{zero}} = \{x \in \text{Supp}_{F_0}(X_i) : \mu(x) = 0\} \neq \emptyset$, and $\mu(\cdot)$ is continuous on R, and (vi) under F_0, (X_i, U_i) are i.i.d., X_i and U_i are independent, $E_{F_0}U_i = 0$, $\text{Var}_{F_0}(U_i) = 1$, X_i is absolutely continuous, and $\text{Var}_{F_0}(X_i) \in (0, \infty)$. As defined, the conditional moment inequality is

$$E_{F_0}\big(m(W_i, \theta_0)|X_i\big) = \mu(X_i) \geq 0 \quad \text{a.s.} \tag{13.26}$$

The inequality in (13.26) is strict except when $X_i \in \mathcal{X}_{\text{zero}}$. Often, the latter occurs with probability zero. For example, this is true if $\mathcal{X}_{\text{zero}}$ is a singleton (or a set with Lebesgue measure zero). In spite of the moment inequality being strict with probability 1, the true value θ_0 is on the boundary of the identified set Θ_{F_0}, that is, $\Theta_{F_0} = (-\infty, \theta_0]$.[56]

We consider a test statistic based on $S(n^{1/2}\bar{m}_n(\theta, g), I)$ with $S = S_1 = S_2$:

$$\begin{aligned} T_n(\theta_0) &= \int \big[n^{1/2}\bar{m}_n(\theta_0, g)\big]_-^2\, dQ(g) \\ &= \int \Bigg[n^{1/2}\Bigg(n^{-1}\sum_{i=1}^n \big(U_i + \mu(X_i)\big)g(X_i) - \Delta(g)\Bigg) \\ &\qquad + n^{1/2}\Delta(g)\Bigg]_-^2 dQ(g), \quad \text{where} \end{aligned} \tag{13.27}$$

$$\bar{m}_n(\theta_0, g) = n^{-1}\sum_{i=1}^n (Y_i - \theta_0)g(X_i) \quad \text{and} \quad \Delta(g) = E_{F_0}\mu(X_i)g(X_i).$$

The first summand in the integrand in (13.27) is $O_p(1)$ uniformly over $g \in \mathcal{G}$ by a functional central limit theorem (CLT) and is identically zero if $P_{F_0}(g(X_i) = 0) = 1$. The second summand, $n^{1/2}\Delta(g)$, diverges to infinity unless $\Delta(g) = 0$. In addition, $[x_n]_-^2 \to 0$ as $x_n \to \infty$. Hence, if $\Delta(g) > 0$, the integrand converges

[56] This holds because, for any $\theta > \theta_0$, (a) $E_{F_0}(m(W_i, \theta)|X_i) = \mu(X_i) + \theta_0 - \theta$, (b) $\forall \delta > 0$, $P_{F_0}(X_i \in B(\mathcal{X}_{\text{zero}}, \delta)) > 0$ by the absolute continuity of X_i, where $B(\mathcal{X}_{\text{zero}}, \delta)$ denotes the closed set of points that are within δ of the set $\mathcal{X}_{\text{zero}}$, (c) for $\delta^* > 0$ sufficiently small, $\mu(x) < \theta - \theta_0\ \forall x \in B(\mathcal{X}_{\text{zero}}, \delta^*)$ by the continuity of $\mu(\cdot)$, and, hence, (d) $0 < P_{F_0}(X_i \in B(\mathcal{X}_{\text{zero}}, \delta^*)) \leq P_{F_0}(E_{F_0}(m(W_i, \theta)|X_i) < 0)$, which implies that $\theta \notin \Theta_{F_0}$.

in probability to zero. In the leading case in which $\mathcal{X}_{\text{zero}}$ is a singleton set (or any set with Lebesgue measure zero), $\Delta(g)=0$ only if $P_{F_0}(g(X_i)=0)=1$ (using the absolute continuity of X_i). In consequence, if $\Delta(g)=0$, the integrand in (13.27) equals zero a.s. Combining these results shows that the asymptotic distribution of $T_n(\theta_0)$ under the fixed distribution F_0 is pointmass at zero even though the true parameter is on the boundary of the identified set.[57]

The pointmass asymptotic distribution of $T_n(\theta_0)$ does not mimic its finite-sample distribution well at all. In finite samples, the distribution of $T_n(\theta_0)$ is nondegenerate because the quantity $n^{1/2}\Delta(g)$ is finite and far from infinity for all functions g for which $\mu(x)$ is not large for $x\in \operatorname{Supp}(g)$. Pointwise asymptotics fail to capture this.

The implication of the discussion above is that, to obtain asymptotic results that mimic the finite-sample situation, it is useful to consider uniform asymptotics or, at least, asymptotics under drifting sequences of distributions.

13.7. *Subsampling Critical Values*

13.7.1. *Definition*

Here we define subsampling critical values and CS's. Let b denote the subsample size when the full sample size is n. We assume $b\to\infty$ and $b/n\to 0$ as $n\to\infty$. The number of different subsamples of size b is q_n. There are $q_n=n!/(b!(n-b)!)$ different subsamples of size b.

Let $\{T_{n,b,j}(\theta): j=1,\dots,q_n\}$ be subsample statistics, where $T_{n,b,j}(\theta)$ is defined exactly the same as $T_n(\theta)$ is defined but based on the jth subsample rather than the full sample. The empirical distribution function and the $1-\alpha$ quantile of $\{T_{n,b,j}(\theta): j=1,\dots,q_n\}$ are

$$(13.28)\quad U_{n,b}(\theta,x)=q_n^{-1}\sum_{j=1}^{q_n}1\big(T_{n,b,j}(\theta)\le x\big)\quad\text{for}\quad x\in R\quad\text{and}$$
$$c_{n,b}(\theta,1-\alpha)=\inf\big\{x\in R: U_{n,b}(\theta,x)\ge 1-\alpha\big\},$$

respectively. The subsampling critical value is $c_{n,b}(\theta_0,1-\alpha)$. The nominal level $1-\alpha$ CS is given by (2.5) with $c_{n,1-\alpha}(\theta)=c_{n,b}(\theta,1-\alpha)$.[58]

[57]This argument is only heuristic. The result can be proved formally using a combination of an almost sure representation result and the bounded convergence theorem, as in the proofs given in Supplemental Appendix A.

[58]The subsampling critical value defined above is a non-recentered subsampling critical value. One also could consider recentered subsampling critical values; see Andrews and Soares (2010) for the definition. But, there is little reason to do so because tests based on recentered subsampling critical values have the same first-order asymptotic power properties as PA tests and recentered bootstrap tests and worse behavior than the latter two tests in terms of the magnitude of errors in null rejection probabilities asymptotically.

13.7.2. *Asymptotic Coverage Probabilities of Subsampling Confidence Sets*

Next, we show that nominal $1-\alpha$ subsampling CS's have asymptotic coverage probabilities greater than or equal to $1-\alpha$ under drifting sequences of parameters and distributions $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$. The sequences that we consider are those in the set Seq^b, which is defined as follows.

Let $\mathcal{H}_1, \mathcal{H}_2$, and $\mathcal{H}$ be defined as in (5.5). Let $\mathcal{H}_1^*(h_1) = \{h_1^* \in \mathcal{H}_1 : h_{1,j}^*(g) > 0$ only if $h_{1,j}(g) = \infty$ for $j \leq p, \forall g \in \mathcal{G}\}$.

DEFINITION $\mathrm{Seq}^b(h_1^*, h)$: For $h \in \mathcal{H}$ and $h_1^* \in \mathcal{H}_1^*(h_1)$, define $\mathrm{Seq}^b(h_1^*, h)$ to be the set of sequences $\{(\theta_n, F_n) : n \geq 1\}$ such that

(i) $(\theta_n, F_n) \in \mathcal{F}\ \forall n \geq 1$,
(ii) $\lim_{n\to\infty} h_{1,n,F_n}(\theta_n, g) = h_1(g)\ \forall g \in \mathcal{G}$,
(iii) $\lim_{n\to\infty} \sup_{g,g^* \in \mathcal{G}} \|D_{F_n}^{-1/2}(\theta_n)\Sigma_{F_n}(\theta_n, g, g^*)D_{F_n}^{-1/2}(\theta_n) - h_2(g, g^*)\| = 0$, and
(iv) $\lim_{n\to\infty} b^{1/2} D_{F_n}^{-1/2}(\theta_n) E_{F_n} m(W, \theta_n, g) = h_1^*(g)\ \forall g \in \mathcal{G}$.

Let

$$\mathrm{Seq}^b = \bigcup_{h_1^* \in \mathcal{H}_1^*(h), h \in \mathcal{H}} \mathrm{Seq}^b(h_1^*, h). \tag{13.29}$$

We use the following assumptions.

ASSUMPTION SQ: *For all functions $h_1 : \mathcal{G} \to R_{[+\infty]}^p \times \{0\}^v$, $h_2 : \mathcal{G}^2 \to \mathcal{W}$, and mean zero Gaussian processes $\{\nu_{h_2}(g) : g \in \mathcal{G}\}$ with finite-dimensional covariance matrix $h_2(g, g^*)$ for $g, g^* \in \mathcal{G}$, the distribution function of $\int S(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k)\, dQ(g)$ at $x \in R$ is*

(a) *continuous for $x > 0$ and*
(b) *strictly increasing for $x > 0$ unless $v = 0$ and $h_1(g) = \infty^p$ a.s. $[Q]$.*

Lemma B4 below shows that Assumption SQ is satisfied by S_1 and S_2.

LEMMA B4: *Assumption SQ holds when $S = S_1$ or S_2.*

The following Assumption C is needed only to show that subsampling CS's are not asymptotically conservative. For $(\theta, F) \in \mathcal{F}$, define $h_{1,j,F}(\theta, g) = \infty$ if $E_F m_j(W_i, \theta, g) > 0$ and $h_{1,j,F}(\theta, g) = 0$ if $E_F m_j(W_i, \theta, g) = 0$ for $g \in \mathcal{G}$, $j = 1, \ldots, p$. Let $h_{1,F}(\theta, g) = (h_{1,1,F}(\theta, g), \ldots, h_{1,p,F}(\theta, g), 0_v')'$.

ASSUMPTION C: *For some $(\theta, F) \in \mathcal{F}$, $\int S(\nu_{h_{2,F}}(\theta, g) + h_{1,F}(\theta, g), h_{2,F}(\theta, g) + \varepsilon I_k)\, dQ(g)$ is continuous at its $1-\alpha$ quantile, where $\{\nu_{h_{2,F}}(\theta, g) : g \in \mathcal{G}\}$ is a mean zero Gaussian process concentrated on the space of uniformly ρ-continuous bounded R^k-valued functionals on $\mathcal{G}$, that is, $U_\rho^k(\mathcal{G})$, with covariance kernel $h_{2,F}(\theta, g, g^*)$ for $g, g^* \in \mathcal{G}$.*

Assumption C is not very restrictive.

The exact and asymptotic confidence sizes of a subsampling CS are

$$\text{(13.30)} \quad \mathrm{ExCS}_n = \inf_{(\theta,F)\in\mathcal{F}} P_F\big(T_n(\theta) \le c_{n,b}(\theta, 1-\alpha)\big) \quad \text{and}$$

$$\mathrm{AsyCS} = \liminf_{n\to\infty} \mathrm{ExCS}_n.$$

The next assumption is used to establish AsyCS for subsampling CS's. It is a high-level condition that is difficult to verify and hence is not very satisfactory.

ASSUMPTION SUB: *For some subsequence* $\{v_n : n \ge 1\}$ *of* $\{n\}$ *for which* $\{(\theta_{v_n}, F_{v_n}) \in \mathcal{F} : n \ge 1\}$ *satisfies* $\lim_{n\to\infty} P_{F_{v_n}}(T_n(\theta_{v_n}) \le c_{n,b}(\theta_{v_n}, 1-\alpha)) = \mathrm{AsyCS}$ (*such a subsequence always exists*), *there is a subsequence* $\{m_n\}$ *of* $\{v_n\}$ *such that* $\{(\theta_{m_n}, F_{m_n}) \in \mathcal{F} : n \ge 1\}$ *belongs to* Seq^b, *where* Seq^b *is defined with* m_n *in place of* n *throughout*.

Part (a) of the following theorem shows that subsampling CS's have correct asymptotic coverage probabilities under drifting sequences of parameters and distributions.

THEOREM B6: *Suppose Assumptions* M, S1, S2, *and* SQ *hold*. *Then*, *a nominal* $1-\alpha$ *subsampling confidence set based on* $T_n(\theta)$ *satisfies*

(a) $\inf_{\{(\theta_n,F_n):n\ge1\}\in\mathrm{Seq}^b} \liminf_{n\to\infty} P_{F_n}(T_n(\theta_n) \le c_{n,b}(\theta_n, 1-\alpha)) \ge 1-\alpha$,

(b) *if Assumption* C *also holds*, *then*

$$\inf_{\{(\theta_n,F_n):n\ge1\}\in\mathrm{Seq}^b} \liminf_{n\to\infty} P_{F_n}\big(T_n(\theta_n) \le c_{n,b}(\theta_n, 1-\alpha)\big) = 1-\alpha,$$

and

(c) *if Assumptions* Sub *and* C *also hold*, *then* $\mathrm{AsyCS} = 1-\alpha$.

COMMENT: Theorem B6(c) establishes that subsampling CS's have correct AsyCS provided Assumption Sub holds. The latter condition is difficult to verify. Hence, this result is not nearly as useful as the uniformity results given for GMS and PA CS's in Section 5.

14. SUPPLEMENTAL APPENDIX C

In this appendix, we prove all the results stated in the main paper except for Theorems 1 and 2(a), which are proved in Supplemental Appendix A, and Lemma A1, which is proved in Supplemental Appendix E. The proofs are given in the following order: Lemma 2, Lemma 3, Theorem 2(b), Lemma 4, Theorem 3, Theorem 4, and Lemma 1.

14.1. *Proofs of Lemmas 2 and 3 and Theorem 2(b)*

PROOF OF LEMMA 2: We have $\theta \notin \Theta_F(\mathcal{G})$ implies that $E_F m_j(W_i, \theta) g_j(X_i) < 0$ for some $j \leq p$ or $E_F m_j(W_i, \theta) g_j(X_i) \neq 0$ for some $j = p+1, \ldots, k$. By the law of iterated expectations and $g_j(x) \geq 0$ for all $x \in R^{d_x}$ and $j \leq p$, this implies that $P_F(X_i \in \mathcal{X}_F(\theta)) > 0$ and, hence, $\theta \notin \Theta_F$.

On the other hand, $\theta \notin \Theta_F$ implies that $P_F(X_i \in \mathcal{X}_F(\theta)) > 0$ and the latter implies that $\theta \notin \Theta_F(\mathcal{G})$ by Assumption CI. *Q.E.D.*

The proof of Lemma 3 uses the following lemma, which is an existence and uniqueness result. The proof of the lemma utilizes an extended measure result from Billingsley (1995, Thm. 11.3), which delivers the existence part of the lemma. The proof is given after the proof of Lemma 3.

LEMMA C1: *Let $\mathcal{R}$ be a semiring of subsets of R^{d_x}. Let μ be a bounded countably additive set function on $\sigma(\mathcal{R})$ such that $\mu(\emptyset) = 0$ and $\mu(C) \geq 0$ for all $C \in \mathcal{R} \cup \{R^{d_x}\}$. If R^{d_x} can be written as the union of a countable number of disjoint sets in $\mathcal{R}$, then μ is a measure on $\sigma(\mathcal{R})$ (and hence $\mu(C) \geq 0$ for all $C \in \sigma(\mathcal{R})$).*[59]

PROOF OF LEMMA 3: First, we establish Assumption CI for $\mathcal{G} = \mathcal{G}_{\text{box}}$ with $\bar{r} = \infty$. It suffices to show that

$$
\begin{aligned}
(14.1)\qquad & E_F\big(m_j(W_i, \theta) g_j(X_i)\big) \geq 0 \quad \forall g \in \mathcal{G} \\
& \quad \Rightarrow \quad E_F\big(m_j(W_i, \theta) | X_i\big) \geq 0 \quad \text{a.s.} \quad \text{for} \quad j = 1, \ldots, p \quad \text{and} \\
& E_F\big(m_j(W_i, \theta) g_j(X_i)\big) = 0 \quad \forall g \in \mathcal{G} \\
& \quad \Rightarrow \quad E_F\big(m_j(W_i, \theta) | X_i\big) = 0 \quad \text{a.s.} \quad \text{for} \quad j = p+1, \ldots, k.
\end{aligned}
$$

We use the following set function:

$$
(14.2)\qquad \mu_j(C) = \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) 1(X_i \in C) \quad \text{for} \quad C \in \sigma(\mathcal{C}_{\text{box}}) = \mathcal{B}\big(R^{d_x}\big),
$$

where $\sigma(\mathcal{C}_{\text{box}})$ denotes the σ-field generated by $\mathcal{C}_{\text{box}}$, $\mathcal{B}(R^{d_x})$ is the Borel σ-field on R^{d_x}, and $\sigma(\mathcal{C}_{\text{box}}) = \mathcal{B}(R^{d_x})$ is a well-known result. First we show $\mu_j(R^{d_x}) \geq 0$. Let $I_L = (-L, L]^{d_x}$. Then, $I_L \in \mathcal{C}_{\text{box}}$ and $\mu_j(I_L) \geq 0$. We have

$$
\begin{aligned}
(14.3)\qquad 0 \leq \lim_{L \to \infty} \mu_j(I_L) &= \lim_{L \to \infty} \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) 1(X_i \in I_L) \\
&= \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) 1\big(X_i \in R^{d_x}\big) = \mu_j\big(R^{d_x}\big),
\end{aligned}
$$

[59]A class of subsets, $\mathcal{R}$, of a universal set is called a semiring if (a) the empty set $\emptyset \in \mathcal{R}$; (b) $A, B \in \mathcal{R}$ implies $A \cap B \in \mathcal{R}$; (c) if $A, B \in \mathcal{R}$ and $A \subset B$, then there exist disjoint sets $C_1, \ldots, C_N \subset \mathcal{R}$ such that $B - A = \bigcup_{i=1}^N C_i$; see Billingsley (1995, p. 138).

where the second equality holds by the dominated convergence theorem, $\sigma_{F,j}^{-1}(\theta)m_j(w,\theta)1(x \in I_L) \to \sigma_{F,j}^{-1}(\theta)m_j(w,\theta)1(x \in R^{d_x})$ as $L \to \infty$, $|\sigma_{F,j}^{-1}(\theta) \times m_j(w,\theta)1(x \in I_L)| \le \sigma_{F,j}^{-1}(\theta)|m_j(w,\theta)|$ for all w, and $\sigma_{F,j}^{-1}(\theta)E_F|m_j(W_i,\theta)| < \infty$.

Next, we treat the cases $j \le p$ and $j > p$ separately because different techniques are employed. First, we consider $j = 1, \dots, p$. Suppose $E_F m_j(W_i,\theta) \times g_j(X_i) \ge 0$ $\forall g \in \mathcal{G}$. Then, $\mu_j(C) \ge 0$ $\forall C \in \mathcal{C}_{\text{box}}$. We want to show that $E_F m_j(W_i,\theta)1(X_i \in C) \ge 0$ $\forall C \in \mathcal{B}(R^{d_x})$, because this implies that $E_F(m_j(W_i,\theta)|X_i) \ge 0$ a.s. since X_i is Borel measurable.

By Lemma C1, we have $\mu_j(C) \ge 0$ $\forall C \in \sigma(\mathcal{C}_{\text{box}})$ if (a) $\mathcal{C}_{\text{box}}$ is a semiring of subsets of R^{d_x}, (b) μ_j is bounded, (c) μ_j is countably additive, (d) $\mu_j(\emptyset) = 0$, (e) $\mu_j(R^{d_x}) \ge 0$, and (f) R^{d_x} can be written as the union of a countable number of disjoint sets in $\mathcal{C}_{\text{box}}$. It is a well-known result that (a) holds (provided $\emptyset$ is added to $\mathcal{C}_{\text{box}}$). By condition (vi) in (2.3), (b) holds. Condition (c) holds by the dominated convergence theorem. Because $1(X_i \in \emptyset) = 0$, condition (d) holds. By (14.3), condition (e) holds. Condition (f) holds because

$$(14.4) \qquad R^{d_x} = \bigcup_{\{i_1,i_2,\dots,i_k\}\in\mathbb{N}^k} \bigtimes_{j=1}^{k} (i_j, i_j+1],$$

where $\mathbb{N}$ is the set of all natural numbers. Therefore, $\mu_j(C) \ge 0$ $\forall C \in \sigma(\mathcal{C}_{\text{box}}) = \mathcal{B}(R^{d_x})$, that is,

$$(14.5) \qquad E_F m_j(W_i,\theta)1(X_i \in C) \ge 0 \quad \forall C \in \mathcal{B}(R^{d_x}).$$

Next, we consider $j = p+1, \dots, k$. Suppose $E_F m_j(W_i,\theta)g_j(X_i) = 0$ $\forall g \in \mathcal{G}_{\text{box}}$. Then, $\mu_j(C) = 0$ $\forall C \in \mathcal{C}_{\text{box}}$. We want to show that $E_F m_j(W_i,\theta)1(X_i \in C) = 0$ $\forall C \in \mathcal{B}(R^{d_x})$, because this implies that $E_F(m_j(W_i,\theta)|X_i) = 0$ a.s. because X_i is Borel measurable. To do so, we show that $\mathcal{C}_0 = \mathcal{B}(R^{d_x})$, where $\mathcal{C}_0 \equiv \{C \in \mathcal{B}(R^{d_x}) : \mu_j(C) = 0\}$. It suffices to show $\mathcal{B}(R^{d_x}) \subset \mathcal{C}_0$. Because $\mathcal{C}_{\text{box}} \subset \mathcal{C}_0$ and $\sigma(\mathcal{C}_{\text{box}}) = \mathcal{B}(R^{d_x})$, it suffices to show that $\mathcal{C}_0$ is a σ-field. The set $\mathcal{C}_0$ is indeed a σ-field because (a) $R^{d_x} \in \mathcal{C}_0$ by (14.3), (b) if $C \in \mathcal{C}_0$, then $\mu_j(C^c) = \mu_j(R^{d_x}) - \mu_j(C) = 0$, that is, $C^c \in \mathcal{C}_0$, and (c) if $C_1, C_2, \dots$ are disjoint sets in $\mathcal{C}_0$, then $\mu_j(\bigcup_{i=1}^{\infty} C_i) = \sum_{i=1}^{\infty} \mu_j(C_i) = 0$ because μ_j is an additive set function, that is, $\bigcup_{i=1}^{\infty} C_i \in \mathcal{C}_0$. This completes the proof of Assumption CI for $\mathcal{G} = \mathcal{G}_{\text{box}}$ with $\bar{r} = \infty$.

Assumption CI holds for $\mathcal{G} = \mathcal{G}_{\text{box}}$ with $\bar{r} = \infty$ implies that Assumption CI holds for $\mathcal{G} = \mathcal{G}_{\text{box}}$ when $\bar{r} \in (0, \infty)$. The reason is that if some deviation is captured by a big box, it also must be captured by some smaller box contained in the big box (because a big box is a finite disjoint union of smaller boxes).

For $\mathcal{G} = \mathcal{G}_{\text{c-cube}}$, Assumption CI holds by the same argument as for $\mathcal{G}_{\text{box}}$ but with $\mathcal{C}_{\text{c-cube}}$ in place of $\mathcal{C}_{\text{box}}$ provided (i) $\mathcal{C}_{\text{c-cube}} \cup \{\phi\}$ is a semiring of subsets of $[0,1]^{d_x}$, (ii) $[0,1]^{d_x}$ can be written as the union of a countable number of disjoint sets in $\mathcal{C}_{\text{c-cube}}$, and (iii) $\sigma(\mathcal{C}_{\text{c-cube}}) = \mathcal{B}([0,1]^{d_x})$. Condition (i) is straightforward to verify. Condition (ii) is verified by using $\bigcup_{\ell=1}^{2r}((\ell-1)/(2r), \ell/(2r)] = [0,1]$ (since the interval $(0, 1/(2r)]$ is defined specially to include 0) to construct a finite number of d_x-dimensional boxes whose union is $[0,1]^{d_x}$. Condition (iii) holds because every element of $\mathcal{C}_{\text{box}}$ can be written as a countable union of sets in $\mathcal{C}_{\text{c-cube}}$ and $\sigma(\mathcal{C}_{\text{box}}) = \mathcal{B}([0,1]^{d_x})$.

Finally, we establish Assumption M. For $\mathcal{G} = \mathcal{G}_{\text{box}}$, Assumptions M(a) and M(b) hold by taking $G(x) = 1$ $\forall x$ and $\delta_1 = 4/\delta + 3$. Assumption M(c) holds because $\mathcal{C}_{\text{box}}$ forms a Vapnik–Cervonenkis class of sets. Assumption M holds for $\mathcal{G}_{\text{c-cube}}$ because $\mathcal{G}_{\text{c-cube}} \subset \mathcal{G}_{\text{box}}$. *Q.E.D.*

PROOF OF LEMMA C1: Because (i) $\mu : \sigma(\mathcal{R}) \to R$ is a bounded countably additive set function, (ii) $\mu(\emptyset) = 0$, and (iii) $\mu(C) \geq 0$ $\forall C \in \mathcal{R}$, Billingsley's (1995) Theorem 11.3 implies that there exists a measure, μ^*, on $\sigma(\mathcal{R})$ that agrees with μ on $\mathcal{R}$. We want to show that μ^* agrees with μ on $\sigma(\mathcal{R})$. That is, we want to show that $\mathcal{C}_{\text{eq}} = \sigma(\mathcal{R})$, where

$$\mathcal{C}_{\text{eq}} = \{C \in \sigma(\mathcal{R}) : \mu^*(C) = \mu(C)\}. \tag{14.6}$$

It suffices to show that $\sigma(\mathcal{R}) \subseteq \mathcal{C}_{\text{eq}}$ because, by definition, $\sigma(\mathcal{R}) \supseteq \mathcal{C}_{\text{eq}}$. We use Dynkin's π-λ theorem; for example, see Billingsley (1995, p. 33), to establish this.

Because $\mathcal{R}$ is a semiring, $\mathcal{R}$ is a π-system. Now, we show that $\mathcal{C}_{\text{eq}}$ is a λ-system. By definition, the set $\mathcal{C}_{\text{eq}}$ is a λ-system if (a) $R^{d_x} \in \mathcal{C}_{\text{eq}}$, (b) $\forall C_1, C_2 \in \mathcal{C}_{\text{eq}}$ such that $C_1 \subset C_2$, $C_2 - C_1 \in \mathcal{C}_{\text{eq}}$, and (c) $\forall C_1, C_2, \ldots \in \mathcal{C}_{\text{eq}}$ such that $C_i \uparrow C$, $C \in \mathcal{C}_{\text{eq}}$. We show (a), (b), and (c) in turn.

(a) By assumption, R^{d_x} can be written as the union of countable disjoint $\mathcal{R}$-sets, say $C_1, C_2, \ldots \in \mathcal{R}$, where $R^{d_x} = \bigcup_{i=1}^{n} C_i$. By countable additivity of both μ and μ^*, we have

$$\mu(R^{d_x}) = \sum_{i=1}^{\infty} \mu(C_i) = \sum_{i=1}^{\infty} \mu^*(C_i) = \mu^*(R^{d_x}), \tag{14.7}$$

where the second equality holds because $C_1, C_2, \ldots \in \mathcal{R}$ and μ^* agrees with μ on $\mathcal{R}$. Thus condition (a) holds.

(b) Suppose $C_1, C_2 \in \mathcal{C}_{\text{eq}}$ and $C_1 \subset C_2$; then $C_2 = (C_2 - C_1) \cup C_1$. Thus,

$$\mu(C_2 - C_1) = \mu(C_2) - \mu(C_1) = \mu^*(C_2) - \mu^*(C_1) = \mu^*(C_2 - C_1), \tag{14.8}$$

where the first and the third equalities hold by the countable additivity of μ and μ^* and the second equality holds because $C_1, C_2 \in \mathcal{C}_{\text{eq}}$. Thus, condition (b) holds.

(c) Suppose $C_1, C_2, \ldots \in \mathcal{C}_{\text{eq}}$ and $C_i \uparrow C$; then $C = C_1 \cup (\bigcup_{i=2}^{\infty}(C_i - C_{i-1}))$ and $C_1, C_2 - C_1, \ldots$ are mutually disjoint. By condition (b), $C_i - C_{i-1} \in \mathcal{C}_{\text{eq}}$ for $i \geq 2$.

Thus,

$$\begin{aligned}(14.9)\qquad \mu(C) &= \mu(C_1) + \sum_{i=2}^{\infty} \mu(C_i - C_{i-1}) \\ &= \mu^*(C_1) + \sum_{i=2}^{\infty} \mu^*(C_i - C_{i-1}) = \mu^*(C).\end{aligned}$$

That is, condition (c) holds.

Therefore, $\mathcal{C}_{\text{eq}}$ is a λ-system. Because $\mathcal{R} \subset \mathcal{C}_{\text{eq}}$ by Dynkin's π-λ theorem, $\sigma(\mathcal{R}) \subseteq \mathcal{C}_{\text{eq}}$. In consequence, $\sigma(\mathcal{R}) = \mathcal{C}_{\text{eq}}$, that is, μ^* agrees with μ on $\sigma(\mathcal{R})$. Because μ^* is a measure on $\sigma(\mathcal{R})$, μ must be a measure on $\sigma(\mathcal{R})$. *Q.E.D.*

PROOF OF THEOREM 2(b): Consider the parameters (θ_c, F_c) that appear in Assumption GMS2. First, we determine the asymptotic behavior of the critical value $c(\varphi_n(\theta_c), \widehat{h}_{n,2}(\theta_c), 1-\alpha)$ under (θ_c, F_c). We have

$$\begin{aligned}(14.10)\qquad \xi_n(\theta_c, g) &= \kappa_n^{-1} n^{1/2} \bar{D}_n^{-1/2}(\theta_c, g) \bar{m}_n(\theta_c, g) \\ &= \bar{D}_n^{-1/2}(\theta_c, g) D_{F_c}^{1/2}(\theta_c) \kappa_n^{-1} \big[\nu_{n,F_c}(\theta_c, g) + h_{1,n,F_c}(\theta_c, g)\big] \\ &= \mathrm{Diag}^{-1/2}\big(\bar{h}_{2,n,F_c}(\theta_c, g)\big) \kappa_n^{-1} \big[\nu_{n,F_c}(\theta_c, g) + h_{1,n,F_c}(\theta_c, g)\big].\end{aligned}$$

Note that $\bar{h}_{2,n,F_c}(\theta_c, g)$ is a function of $\widehat{h}_{2,n,F_c}(\theta_c, g, g)$ by (5.2). Let

$$(14.11)\qquad T_n^{\text{GMS}}(\theta_c) = \int S\big(\nu_{\widehat{h}_{2,n}(\theta_c)}(g) + \varphi_n(\theta_c, g), \widehat{h}_{2,n}(\theta_c, g) + \varepsilon I_k\big)\, dQ(g),$$

where $\{\nu_{h_2}(g) : g \in \mathcal{G}\}$ is defined as in (4.2) on the same probability space as the observations and is independent of the observations, and $\nu_{\widehat{h}_{2,n}(\theta_c)}(\cdot)$ equals $\nu_{h_2}(\cdot)$ evaluated at $h_2 = \widehat{h}_{2,n}(\theta_c)$. Equations (4.10), (12.26), (14.10), and (14.11) imply that the distribution of $T_n^{\text{GMS}}(\theta_c)$ is determined by the joint distribution of $\{\nu_{\widehat{h}_{2,n}(\theta_c)}(g) : g \in \mathcal{G}\}$, $\{\widehat{h}_{2,n,F_c}(\theta_c, g) : g \in \mathcal{G}\}$, and $\{\kappa_n^{-1}\nu_{n,F_c}(\theta_c, g) : g \in \mathcal{G}\}$.

We have $\{(\theta_c, F_c) : n \geq 1\} \in \mathrm{SubSeq}(h_{2,F_c}(\theta_c))$ because $(\theta_c, F_c) \in \mathcal{F}$. Hence, by Lemma A1(b), $d(\widehat{h}_{2,n,F_c}(\theta_c), h_{2,F_c}(\theta_c)) \to_p 0$ as $n \to \infty$. By the same argument as in (12.26), this yields $d(\widehat{h}_{2,n}(\theta_c), h_{2,F_c}(\theta_c)) \to_p 0$. The latter, the independence of $\widehat{h}_{2,n,F_c}(\theta_c)$ and $\{\nu_{h_2}(\cdot) : h_2 \in \mathcal{H}_2\}$, and an almost sure representation argument imply that the Gaussian processes $\{\nu_{\widehat{h}_{2,n}(\theta_c)}(\cdot) : n \geq 1\}$ converge weakly to $\nu_{h_{2,F_c}(\theta_c)}(\cdot)$ as $n \to \infty$. The sequence of random processes $\{\widehat{h}_{2,n}(\theta_c, \cdot) : n \geq 1\}$ converges in probability uniformly (and hence in distribution) to $h_{2,F_c}(\theta_c, \cdot)$, where $\widehat{h}_{2,n}(\theta_c, g) = \widehat{h}_{2,n}(\theta_c, g, g)$ and $h_{2,F_c}(\theta_c, g) = h_{2,F_c}(\theta_c, g, g)$. The sequence $\{\kappa_n^{-1}\nu_{n,F_c}(\theta_c, \cdot) : n \geq 1\}$ converges in probability to zero uniformly over $g \in \mathcal{G}$ because $\kappa_n \to \infty$ and $\{\nu_{n,F_n}(\theta_c, \cdot) : n \geq 1\}$ converges to a Gaussian process with sample paths that are bounded a.s. Therefore, we have

$$(14.12)\quad \begin{pmatrix} \nu_{\widehat{h}_{2,n}(\theta_c)}(\cdot) \\ \widehat{h}_{2,n}(\theta_c,\cdot) \\ \kappa_n^{-1}\nu_{n,F_c}(\theta_c,\cdot) \end{pmatrix} \Rightarrow \begin{pmatrix} \nu_{h_{2,F_c}(\theta_c)}(\cdot) \\ h_{2,F_c}(\theta_c,\cdot) \\ 0_{\mathcal{G}} \end{pmatrix} \quad \text{as} \quad n\to\infty,$$

where $\widehat{h}_{2,n}(\theta_c)$ that appears in $\nu_{\widehat{h}_{2,n}(\theta_c)}(\cdot)$ is a function on $\mathcal{G}\times\mathcal{G}$ whereas $\widehat{h}_{2,n}(\theta_c,\cdot)$ is a function on $\mathcal{G}$, likewise for $\nu_{h_{2,F_c}(\theta_c)}(\cdot)$ and $h_{2,F_c}(\theta_c,\cdot)$, and $0_{\mathcal{G}}$ denotes the R^k-valued function on $\mathcal{G}$ that is identically $(0,\dots,0)'\in R^k$.

By the almost sure representation theorem (see Pollard (1990, Thm. 9.4)), there exist $\{(\tilde{\nu}_n(g),\tilde{h}_{2,n}(g),\ \tilde{\nu}_{\kappa,n}(g)) : g\in\mathcal{G}, n\ge 1\}$ and $\{\tilde{\nu}(g),\tilde{h}_2(g) : g\in\mathcal{G}\}$ such that (i) $\{(\tilde{\nu}_n(g),\tilde{h}_{2,n}(g),\tilde{\nu}_{\kappa,n}(g)) : g\in\mathcal{G}\}$ has the same distribution as $\{(\nu_{\widehat{h}_{2,n}(\theta_c)}(g),\ \widehat{h}_{2,n}(\theta_c,g),\ \kappa_n^{-1}\nu_{n,F_c}(\theta_c,g)) : g\in\mathcal{G}\}$ for all $n\ge 1$, (ii) $\{(\tilde{\nu}(g),$ $\tilde{h}_2(g)) : g\in\mathcal{G}\}$ has the same distribution as $\{(\nu_{h_{2,F_c}(\theta_c)}(g), h_{2,F_c}(\theta_c,g) : g\in\mathcal{G}\}$, and

$$(14.13)\quad \text{(iii)}\quad \sup_{g\in\mathcal{G}}\left\| \begin{pmatrix} \tilde{\nu}_n(g) \\ \tilde{h}_{2,n}(g) \\ \tilde{\nu}_{\kappa,n}(g) \end{pmatrix} - \begin{pmatrix} \tilde{\nu}(g) \\ \tilde{h}_2(g) \\ 0 \end{pmatrix} \right\| \to 0 \quad \text{a.s.}$$

Let

$$(14.14)\quad \tilde{T}_n^{\mathrm{GMS}} = \int S\big(\tilde{\nu}_n(g)+\widetilde{\varphi}_n(g), \tilde{h}_{2,n}(g)+\varepsilon I_k\big)\, dQ(g),$$

where $\widetilde{\varphi}_n(g)$ is defined just as $\varphi_n(\theta,g)$ is defined in (4.10) but with $\tilde{h}_{2,n,j}(g)+\varepsilon\tilde{h}_{2,n,j}(1_k)$ in place of $\bar{h}_{2,n,F_n,j}(\theta,g)$, where $\tilde{h}_{2,n,j}(g)$ denotes the (j,j) element of $\tilde{h}_{2,n}(g)$, and $\tilde{\xi}_n(g)$ in place of $\xi_n(\theta,g)$, where

$$(14.15)\quad \tilde{\xi}_n(g) = \mathrm{Diag}\big(\tilde{h}_{2,n}(g)+\varepsilon\tilde{h}_{2,n}(1_k)\big)^{-1/2}\big(\kappa_n^{-1}\tilde{\nu}_{\kappa,n}(g)+\kappa_n^{-1}h_{1,n,F_c}(\theta_c,g)\big).$$

Then, $\tilde{T}_n^{\mathrm{GMS}}$ and $T_n^{\mathrm{GMS}}(\theta_c)$ have the same distribution for all $n\ge 1$ and the same asymptotic distribution as $n\to\infty$. Let $\tilde{c}_n(1-\alpha)$ denote the $1-\alpha+\eta$ quantile of $\tilde{T}_n^{\mathrm{GMS}}$ plus η, where η is as in the definition of $c(h,1-\alpha)$. Then, $\tilde{c}_n(1-\alpha)$ has the same distribution as $c(\varphi_n(\theta_c),\widehat{h}_{2,n}(\theta_c),1-\alpha)$ for all $n\ge 1$.

Let $\tilde{\Omega}^*$ be the collection of $\omega\in\Omega$ such that, at ω, $\tilde{\nu}(g)(\omega)$ is bounded and the convergence in (14.13) holds. By (14.13) and the fact that the sample paths of $\{\tilde{\nu}(g) : g\in\mathcal{G}\}$ are bounded a.s., we have $P_{F_c}(\tilde{\Omega}^*)=1$.

Under (θ_c,F_c) for all $n\ge 1$,

$$(14.16)\quad \kappa_n^{-1}h_{1,n,F_c}(\theta_c,g) = \kappa_n^{-1}n^{1/2}D_{F_c}^{-1/2}(\theta_c)E_{F_c}m(W_i,\theta_c,g) \to h_{1,\infty,F_c}(\theta_c,g)$$

as $n \to \infty$, using Assumption GMS2(c). Thus, for fixed $\omega \in \tilde{\Omega}^*$,

$$\begin{aligned}(14.17)\quad \tilde{\xi}_n(g)(\omega) &= \mathrm{Diag}^{-1/2}\big(\tilde{h}_2(g) + \varepsilon \tilde{h}_2(1_k) + o(1)\big) \\ &\quad \times \big(o(1) + \kappa_n^{-1} h_{1,n,F_c}(\theta_c, g)\big) \\ &\to h_{1,\infty,F_c}(\theta_c, g),\end{aligned}$$

as $n \to \infty$ for all $g \in \mathcal{G}$, where $\tilde{h}_{2,j}(g)$ denotes the (j, j) element of $\tilde{h}_2(g)$, using (14.13), $\tilde{h}_2(1_k) = I_k$ (which holds by (5.1) and Definition SubSeq(h_2)), $\tilde{h}_{2,j}(g) \geq 0$, $\varepsilon > 0$.

By (14.17), Assumption GMS1(a), $B_n \to \infty$ as $n \to \infty$ (by Assumption GMS2(b)), and the fact that $h_{1,\infty,F_c}(\theta_c, g)$ equals either 0 or ∞ by definition, we have

$$(14.18)\quad \widetilde{\varphi}_n(g)(\omega) \to h_{1,\infty,F_c}(\theta_c, g) \quad \text{as} \quad n \to \infty$$

for all $\omega \in \tilde{\Omega}^*$.

By (14.13), (14.15), (14.18), and Assumption S1(d), we have

$$\begin{aligned}(14.19)\quad & S\big(\tilde{\nu}_n(g) + \widetilde{\varphi}_n(g), \tilde{h}^*_{2,n}(g) + \varepsilon I_k\big)(\omega) \\ &\quad \to S\big(\tilde{\nu}(g) + h_{1,\infty,F_c}(\theta_c, g), h_{2,F_c}(\theta_c, g) + \varepsilon I_k\big)(\omega)\end{aligned}$$

as $n \to \infty$ $\forall \omega \in \tilde{\Omega}^*, \forall g \in \mathcal{G}$. Now, by the argument given from (12.14) to the end of the proof of Theorem 1, the quantity on the left-hand side of (14.19) is bounded by a finite constant. This, (14.19), and the bounded convergence theorem give

$$(14.20)\quad \tilde{T}_n^{\mathrm{GMS}} \to \tilde{T}^{\mathrm{GMS}} = \int S\big(\tilde{\nu}(g) + h_{1,\infty,F_c}(\theta_c, g), h_{2,F_c}(\theta_c, g) + \varepsilon I_k\big)\, dQ(g)$$

as $n \to \infty$ a.s.

By (14.20),

$$(14.21)\quad P\big(\tilde{T}_n^{\mathrm{GMS}} \leq x\big) \to P\big(\tilde{T}^{\mathrm{GMS}} \leq x\big) \quad \text{as} \quad n \to \infty$$

for all continuity points x of the distribution of $\tilde{T}^{\mathrm{GMS}}$. Let $\tilde{c}_0(1-\alpha)$ denote the $1-\alpha$ quantile of $\tilde{T}^{\mathrm{GMS}}$. Let $\tilde{c}(1-\alpha) = \tilde{c}_0(1-\alpha+\eta) + \eta$, where η is as in the definition of $c(h, 1-\alpha)$. By Assumption GMS2(a), the distribution function of $\tilde{T}^{\mathrm{GMS}}$, which equals that of $T(h_{\infty,F_c}(\theta_c))$, is continuous and strictly increasing at $x = \tilde{c}(1-\alpha)$. Using Lemma 5 of Andrews and Guggenberger (2010), this gives

$$\begin{aligned}(14.22)\quad & \tilde{c}_n(1-\alpha) \to_p \tilde{c}(1-\alpha) \quad \text{and} \\ & c\big(\varphi_n(\theta_c), \widehat{h}_{2,n}(\theta_c), 1-\alpha\big) \to_p \tilde{c}(1-\alpha),\end{aligned}$$

where the second convergence result holds because $\tilde{c}_n(1-\alpha)$ and $c(\varphi_n(\theta_c), \widehat{h}_{2,n}(\theta_c), 1-\alpha)$ have the same distribution.

Next, by the same argument as used above to show (14.20), but with $\nu_{\widehat{h}_{2,n}(\theta_c)}(g)$ and $\varphi_n(\theta_c, g)$ replaced by $\nu_{n,F_c}(\theta_c, g)$ and $h_{1,n,F_c}(\theta_c, g)$, respectively, we have

$$
\begin{aligned}
(14.23)\quad T_n(\theta_c) &\to_d T\big(h_{\infty,F_c}(\theta_c)\big) \\
&= \int S\big(\nu_{h_{2,F_c}(\theta_c)}(g) + h_{1,\infty,F_c}(\theta_c, g), \\
&\qquad h_{2,F_c}(\theta_c, g) + \varepsilon I_k\big)\, dQ(g),
\end{aligned}
$$

where $h_{\infty,F_c}(\theta_c) = (h_{1,\infty,F_c}(\theta_c), h_{2,F_c}(\theta_c))$, $h_{1,n,F_c}(\theta_c) \to h_{1,\infty,F_c}(\theta_c)$ by straightforward calculations, and $\nu_{n,F_c}(\theta_c, \cdot) \Rightarrow \nu_{h_{2,F_c}(\theta_c)}(\cdot)$ by Lemma A1(a). Note that $T(h_{\infty,F_c}(\theta_c))$ and $\tilde{T}^{\mathrm{GMS}}$ have the same distribution because $\nu_{h_{2,F_c}(\theta_c)}(\cdot)$ and $\tilde{\nu}(\cdot)$ have the same distribution. Thus, $\tilde{c}(1-\alpha)$ $(= \tilde{c}_0(1-\alpha+\eta)+\eta)$ is the $1-\alpha+\eta$ quantile of $T(h_{\infty,F_c}(\theta_c))$ plus η.

By (14.22), (14.23), Assumption GMS2(a), and Lemma 5 of Andrews and Guggenberger (2010), for $\eta > 0$, we have

$$
\begin{aligned}
(14.24)\quad \lim_{n\to\infty} P_{F_c}\big(T_n(\theta_c) \le c\big(\varphi_n(\theta_c), \widehat{h}_{2,n}(\theta_c), 1-\alpha\big)\big) \\
= P\big(T\big(h_{\infty,F_c}(\theta_c)\big) \le \tilde{c}_0(1-\alpha+\eta)+\eta\big).
\end{aligned}
$$

The limit as $\eta \to 0$ of the right-hand side equals $1-\alpha$ because distribution functions are right-continuous and the distribution function of $T(h_{\infty,F_c}(\theta_c))$ at its $1-\alpha$ quantile equals $1-\alpha$ by Assumption GMS2(a).

Combining (14.24) and the result of Theorem 2(a), which holds for all $\eta > 0$ and hence holds when the limit as $\eta \to 0$ is taken, gives Theorem 2(b). *Q.E.D.*

14.2. *Proofs of Results for Fixed Alternatives*

PROOF OF LEMMA 4: First, we prove part (a). It holds immediately that $\mathrm{Supp}(Q_a) = \mathcal{G}_{\text{c-cube}}$ because $\mathcal{G}_{\text{c-cube}}$ is countable and Q_a has a probability mass function that is positive at each element in $\mathcal{G}_{\text{c-cube}}$.

Next, for part (b), consider $g = g_{x,r} \in \mathcal{G}_{\text{box}}$, where $g_{x,r}(y) = 1(y \in C_{x,r}) \cdot 1_k$ and $(x, r) \in [0,1]^{d_x} \times (0, \bar{r})^{d_x}$. Let $\delta > 0$ be given. The idea of the proof is to find a set $G_{g,\bar{\eta}} \subset \mathcal{B}_{\rho_X}(g, \delta)$ $(\subset \mathcal{G}_{\text{box}})$ such that $Q_b(G_{g,\bar{\eta}}) > 0$. This implies that $Q_b(\mathcal{B}_{\rho_X}(g, \delta)) > 0$, which is the desired result.

The set $G_{g,\bar{\eta}}$ needs to be defined differently (for reasons stated below) depending on whether $x_u < 1$ or $x_u = 1$, for $u = 1, \ldots, d_x$, where $x =$

$(x_1, \ldots, x_{d_x})'$. For $\bar{\eta} > 0$, define

$$
\begin{aligned}
(14.25)\quad G_{g,\bar{\eta}} &= \left\{ g_{x+\eta_0, r+\eta_1} : (\eta_0, \eta_1) \in \Xi_{g,\bar{\eta}} \right\}, \quad \text{where} \\
\Xi_{g,\bar{\eta}} &= \left\{ (\eta_0, \eta_1) \in R^{2d_x} : \text{for } u = 1, \ldots, d_x, \right. \\
&\qquad \text{if } x_u < 1,\ \eta_{0,u} \in [\bar{\eta}, 2\bar{\eta}]\ \&\ \eta_{1,u} \in [0, \bar{\eta}] \text{ and} \\
&\qquad \left. \text{for } x_u = 1,\ \eta_{0,u} \in [-\bar{\eta}, 0]\ \&\ \eta_{1,u} \in [-2\bar{\eta}, -\bar{\eta}] \right\}.
\end{aligned}
$$

We have $Q_b(G_{g,\bar{\eta}}) = Q_b^*((x, r) + \Xi_{g,\bar{\eta}}) > 0$ for all $\bar{\eta} > 0$ because Q_b^* is the uniform distribution on $[0, 1]^{d_x} \times (0, \bar{r})^{d_x}$.

Next, we show that $G_{g,\bar{\eta}} \subset \mathcal{B}_{\rho_X}(g, \delta)$. Let $U_{(x_u<1)} \subset \{1, \ldots, d_x\}$ be the set of indices u such that $x_u < 1$ and let $U_{(x_u=1)} \subset \{1, \ldots, d_x\}$ be the set of indices u such that $x_u = 1$. Let $g_{x+\eta_0,r+\eta_1} \in G_{g,\bar{\eta}}$. The uth lower endpoints of the $C_{x,r}$ and $C_{x+\eta_0,r+\eta_1}$ boxes are $x_u - r_u$ and $x_u + \eta_{0,u} - (r_u + \eta_{1,u})$, respectively. The lower endpoint of the $C_{x+\eta_0,r+\eta_1}$ box is larger than that of the $C_{x,r}$ box because $\eta_{0,u} - \eta_{1,u} \in [0, 2\bar{\eta}]$ (whether $u \in U_{(x_u<1)}$ or $u \in U_{(x_u=1)}$). The uth upper endpoints of the $C_{x,r}$ and $C_{x+\eta_0,r+\eta_1}$ boxes are $x_u + r_u$ and $x_u + \eta_{0,u} + r_u + \eta_{1,u}$, respectively. If $u \in U_{x_u<1}$, the upper endpoint of the $C_{x+\eta_0,r+\eta_1}$ box is larger than that of the $C_{x,r}$ box because $\eta_{0,u} + \eta_{1,u} \in [0, 3\bar{\eta}]$. If $u \in U_{(x_u=1)}$, the uth upper endpoint of the $C_{x+\eta_0,r+\eta_1}$ box is smaller than that of the $C_{x,r}$ box because $\eta_{0,u} + \eta_{1,u} \in [-3\bar{\eta}, 0]$.

Using the results of the previous paragraph, we have

$$
\begin{aligned}
(14.26)\quad & \rho_X^2(g_{x,r}, g_{x+\eta_0,r+\eta_1}) \\
&= E_{F_{X,0}} \left[1(X_i \in C_{x,r}) - 1(X_i \in C_{x+\eta_0,r+\eta_1}) \right]^2 \\
&\le \sum_{u=1}^{d_x} P_{F_{X,0}} \left(X_{i,u} \in \left(x_u - r_u, x_u + \eta_{0,u} - (r_u + \eta_{1,u}) \right] \right) \\
&\quad + \sum_{u \in U_{(x_u<1)}} P_{F_{X,0}} \left(X_{i,u} \in (x_u + r_u, x_u + \eta_{0,u} + r_u + \eta_{1,u}] \right) \\
&\quad + \sum_{u \in U_{(x_u=1)}} P_{F_{X,0}} \left(X_{i,u} \in (1 + \eta_{0,u} + r_u + \eta_{1,u}, 1 + r_u] \cap [0, 1] \right) \\
&\le \sum_{u=1}^{d_x} P_{F_{X,0}} \left(X_{i,u} \in (x_u - r_u, x_u - r_u + 2\bar{\eta}] \right) \\
&\quad + \sum_{u \in U_{(x_u<1)}} P_{F_{X,0}} \left(X_{i,u} \in (x_u + r_u, x_u + r_u + 3\bar{\eta}] \right) \\
&\quad + \sum_{u \in U_{(x_u=1)}} P_{F_{X,0}} \left(X_{i,u} \in (1 + r_u - 3\bar{\eta}, 1 + r_u] \cap [0, 1] \right),
\end{aligned}
$$

where the first inequality uses the d_x-dimensional extension of the one-dimensional result that $(a,b]\,\Delta\,(c,d] \subset (a,c] \cup (b,d]$ when $a \le c$ and $b \le d$, where Δ denotes the symmetric difference of two sets.

The first and second summands on the right-hand side (r.h.s.) of (14.26) tend to zero as $\bar{\eta} \downarrow 0$ by the right-continuity of distribution functions. The third summand on the r.h.s. equals zero when $\bar{\eta}$ is sufficiently small (i.e., when $3\bar{\eta} < \min_{u \le d_x} r_u$). Therefore, for $\bar{\eta} > 0$ sufficiently small, $\rho_X^2(g_{x,r}, g_{x+\eta_0, r+\eta_1}) < \delta$ and $G_{g,\bar{\eta}} \subset \mathcal{B}_{\rho_X}(g, \delta)$. This completes the proof of part (b).

Note that, in the proof of part (b), we cannot treat the case where $u \in U_{(x_u=1)}$ in the same way that we treat the case for $u \in U_{(x_u<1)}$ because, for $u \in U_{(x_u<1)}$, we use the center point $x_u + \eta_{0,u} > x_u$, which is not in $[0,1]$ if $x_u = 1$ and hence violates the assumption of part (b) that the centers of the $\mathcal{G}_{\text{box}}$ boxes lie in $[0,1]^{d_x}$. Conversely, we cannot treat the case where $u \in U_{(x_u<1)}$ in the same way that we treat the case for $u \in U_{(x_u=1)}$ because doing so would lead to a term $P_{F_{X,0}}(X_{i,u} \in (1 + r_u - 3\bar{\eta}, 1 + r_u])$ in (14.26) that does not go to zero as $\bar{\eta} \downarrow 0$ if $X_{i,u}$ has positive probability of equaling $1 + r_u$. *Q.E.D.*

PROOF OF THEOREM 3: Part (a) follows from part (b) because

$$c\big(\varphi_n(\theta_*), \widehat{h}_{2,n}(\theta_*), 1-\alpha\big) \le c\big(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha\big), \tag{14.27}$$

which holds because $\varphi_n(\theta_*, g) \ge 0_k$ $\forall g \in \mathcal{G}$ by Assumption GMS1(a), $c(h_1, \widehat{h}_{2,n}(\theta_*), 1-\alpha)$ is non-increasing in the first p elements of h_1 by Assumption S1(b), and the last v elements of $\varphi_n(\theta_*, g)$ equal zero.

Now we prove part (b). By Assumptions FA(a) and CI, $\beta(g_0) > 0$ for some $g_0 \in \mathcal{G}$. By construction, $e_j = m_j^*(g_0)/\beta(g_0) \in [-1, \infty)$ for $j = 1, \dots, k$ and $e_j = -1$ for some $j \le p$ or $|e_j| = 1$ for some $j = p+1, \dots, k$. As defined above, $\mathcal{B}_{\rho_X}(g_0, \tau_2)$ denotes a ρ_X-ball centered at g_0 with radius $\tau_2 > 0$, where ρ_X is defined in (6.3). First we show that, for some $\tau_2 > 0$,

$$\int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\big(m^*(g)/\beta(g_0), h_{2,0}(g) + \varepsilon I_k\big)\, dQ(g) > 0, \tag{14.28}$$

where $m^*(g) = (m_1^*(g), \dots, m_k^*(g))'$ and $h_{2,0}(g) = h_{2,F_0}(\theta_*, g)$. We have: for $j = 1, \dots, k$,

$$\begin{aligned} (14.29)\quad & \big|m_j^*(g) - m_j^*(g_0)\big| \\ &= \big|E_{F_0} m_j(W_i, \theta_*) g_j(X_i) - E_{F_0} m_j(W_i, \theta_*) g_{0,j}(X_i)\big| / \sigma_{F_0,j}(\theta_*) \\ &\le \big(E_{F_0} m_j^2(W_i, \theta_*)\big)^{1/2} \big(E_{F_0}\big(g_j(X_i) - g_{0,j}(X_i)\big)^2\big)^{1/2} / \sigma_{F_0,j}(\theta_*) \\ &\le \big(E_{F_0}\big\|m(W_i, \theta_*)\big\|^2\big)^{1/2} \rho_X(g, g_0)/\sigma_{F_0,j}(\theta_*), \end{aligned}$$

where $g_{0,j}(X_i)$ denotes the jth element of $g_0(X_i)$.

Given $\tau_1 \in (0,1)$, let

$$\tau_2 = \tau_1 \beta(g_0) \sigma_{F_0,j}(\theta_*) / \bigl(E_{F_0} \|m(W_i, \theta_*)\|^2\bigr)^{1/2}. \tag{14.30}$$

By (14.29), for all $g \in \mathcal{B}_{\rho_X}(g_0, \tau_2)$,

$$\bigl|m_j^*(g) - m_j^*(g_0)\bigr| \le \tau_1 \beta(g_0) \quad \text{for all } j = 1, \dots, k. \tag{14.31}$$

Hence, for all $g \in \mathcal{B}_{\rho_X}(g_0, \tau_2)$, there exists $j \le k$ such that either

$$\begin{aligned} & j \le p \quad \text{and} \quad m_j^*(g)/\beta(g_0) \le m_j^*(g_0)/\beta(g_0) + \tau_1 = -1 + \tau_1 < 0 \quad \text{or} \\ & j \in \{p+1, \dots, k\} \quad \text{and} \\ & \bigl|m_j^*(g)/\beta(g_0)\bigr| \ge \bigl|m_j^*(g_0)/\beta(g_0)\bigr| - \tau_1 = 1 - \tau_1 > 0, \end{aligned} \tag{14.32}$$

using the triangle inequality.

By (14.32) and Assumption S3, $S(m^*(g)/\beta(g_0), h_{2,0}(g) + \varepsilon I_k) > 0$ for all $g \in \mathcal{B}_{\rho_X}(g_0, \tau_2)$. In addition, by Assumption Q, $Q(\mathcal{B}_{\rho_X}(g_0, \tau_2)) > 0$. These properties combine to give (14.28).

We use (14.28) in the following: for all $\delta > 0$,

$$\begin{aligned} & \bigl(n^{1/2}\beta(g_0)\bigr)^{-\chi} T_n(\theta_*) \\ & \quad = \bigl(n^{1/2}\beta(g_0)\bigr)^{-\chi} \\ & \qquad \times \int_{\mathcal{G}} S\bigl(\nu_{n,F_0}(\theta_*, g) + h_{1,n,F_0}(\theta_*, g), \bar{h}_{2,n,F_0}(\theta_*, g)\bigr)\, dQ(g) \\ & \quad \ge \bigl(n^{1/2}\beta(g_0)\bigr)^{-\chi} \\ & \qquad \times \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\bigl(\nu_{n,F_0}(\theta_*, g) + h_{1,n,F_0}(\theta_*, g), \bar{h}_{2,n,F_0}(\theta_*, g)\bigr)\, dQ(g) \\ & \quad = \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\bigl(\bigl(n^{1/2}\beta(g_0)\bigr)^{-1} \nu_{n,F_0}(\theta_*, g) \\ & \qquad + m^*(g)/\beta(g_0), \bar{h}_{2,n,F_0}(\theta_*, g)\bigr)\, dQ(g) \\ & \quad \to_p \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\bigl(m^*(g)/\beta(g_0), h_{2,0}(g) + \varepsilon I_k\bigr)\, dQ(g) \\ & \quad > 0, \end{aligned} \tag{14.33}$$

where χ is as in Assumption S4, the first equality holds by (5.4), the first inequality holds by Assumption S1(c), the second equality holds by Assumption S4 and the definition of $m_j^*(g)$ in (6.2), the last inequality holds by (14.28), and the convergence holds by the argument given in the following paragraph.

By Lemma A1(a) and the continuous mapping theorem, $\sup_{g\in\mathcal{G}} \|\nu_{n,F_0}(\theta_*, g)\| = O_p(1)$. (Note that Lemma A1 applies for $(\theta_{a_n}, F_{a_n}) = (\theta_*, F_0) \notin \mathcal{F}$ for all $n \geq 1$ because Assumptions FA(b)–(d) imply conditions (ii)–(v) in the definition of SubSeq$(h_{2,F_0}(\theta_*))$.) Also, $(n^{1/2}\beta(g_0))^{-1} = o(1)$, because Assumptions FA and CI imply that $\beta(g_0) > 0$ for some $g_0 \in \mathcal{G}$. Hence, (i) $(n^{1/2}\beta(g_0))^{-1}\nu_{n,F_0}(\theta_*, \cdot) \Rightarrow 0_{\mathcal{G}}$. In addition, (ii) $\sup_{g\in\mathcal{G}} \|\bar{h}_{2,n,F_0}(\theta_*, g) - h_{2,0}(g) - \varepsilon I_k\| \to_p 0$, where $h_{2,0}(g) = h_{2,F_0}(\theta_*, g, g)$, by Lemma A1(b), (12.26), and the definition of $\bar{h}_{2,n,F}(\theta, g)$. As in previous proofs, by the almost sure representation theorem, there exist a probability space and random quantities defined on it with the same distributions as $(n^{1/2}\beta(g_0))^{-1}\nu_{n,F_0}(\theta_*, \cdot)$ and $\bar{h}_{2,n,F_0}(\theta_*, \cdot)$ for $n \geq 1$, such that the convergence in (i) and (ii) holds almost surely for these random quantities. Furthermore, using Assumptions S1(b) and S1(e), the integrand in the last equality in (14.33) is bounded by $\sup_{g\in\mathcal{B}^{\mathrm{cl}}_{\rho_X}(g_0,\tau_2), \nu\in R^k:\|\nu\|\leq\delta_*} S(\nu + m^*(g)/\beta(g_0), (\varepsilon - \delta_{**})I_k) < \infty$ for all $g \in \mathcal{B}_{\rho_X}(g_0, \tau_2)$, for some $\delta_*, \delta_{**} > 0$, for n sufficiently large, where $\mathcal{B}^{\mathrm{cl}}_{\rho_X}(g_0, \tau_2)$ denotes the closure of $\mathcal{B}_{\rho_X}(g_0, \tau_2)$, because a continuous function on a compact set attains its supremum using Assumption S1(d) and using an argument analogous to that in (12.14) to treat the second argument of the function S. Thus, by the bounded convergence theorem, the convergence in (14.33) holds a.s. for the newly constructed random quantities. In consequence, it holds in probability for the original random quantities by the equality in distribution of the original and newly constructed random quantities. This completes the proof of the convergence in (14.33).

Next, we show that, under F_0,

$$c\big(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha\big) = O_p(1). \tag{14.34}$$

This and (14.33) give

$$\begin{aligned}
&P_{F_0}\big(T_n(\theta_*) > c\big(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha\big)\big) \\
&\quad = P_{F_0}\Big(\big(n^{1/2}\beta(g_0)\big)^{-\chi} T_n(\theta_*) \\
&\qquad > \big(n^{1/2}\beta(g_0)\big)^{-\chi} c\big(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha\big)\Big) \\
&\quad \geq P_{F_0}\bigg(\int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\big(m^*(g)/\beta(g_0), h_{2,0}(g) + \varepsilon I_k\big)\, dQ(g) + o_p(1) \\
&\qquad > o_p(1)\bigg) \\
&\quad \to 1
\end{aligned} \tag{14.35}$$

as $n \to \infty$, which establishes the result of the theorem.

It remains to show (14.34). Lemma A5, applied with $h_{2,n} = h_{2,0}$, $\{h^*_{2,n} : n \geq 1\}$ being any sequence of $k \times k$-matrix-valued covariance kernels on $\mathcal{G} \times \mathcal{G}$ such that $d(h_{2,n}^*, h_{2,0}) \to 0$, $h_1 = 0_{\mathcal{G}}$, η as in the definition of $c(h, 1-\alpha)$, α replaced by $\alpha - \eta > 0$, and $\eta_1 = \delta$, gives: $\forall \delta > 0$,

$$\begin{aligned}(14.36)\quad & \liminf_{n\to\infty}\big[c_0(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta) \\ & \quad + \delta - c_0\big(0_{\mathcal{G}}, h_{2,n}^*, 1-\alpha+\eta\big)\big] \geq 0 \quad \text{and hence} \\ & \limsup_{n\to\infty} c_0\big(0_{\mathcal{G}}, h_{2,n}^*, 1-\alpha+\eta\big) \leq c_0(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta) + \delta < \infty.\end{aligned}$$

By Lemma A1(b) and (12.26), we obtain $d(\widehat{h}_{2,n}(\theta_*), h_{2,0}) \to_p 0$. As in previous proofs, by the almost sure representation theorem, there exist a probability space and random quantities $\tilde{h}_{2,n}(\cdot,\cdot)$ defined on it with the same distributions as $\widehat{h}_{2,n}(\theta_*,\cdot,\cdot)$ for $n \geq 1$ such that $d(\tilde{h}_{2,n}, h_{2,0}) \to 0$ a.s. This and (14.36) give $\limsup_{n\to\infty} c_0(0_{\mathcal{G}}, \tilde{h}_{2,n}, 1-\alpha+\eta) < \infty$ a.s., which implies (14.34) because $\tilde{h}_{2,n}(\cdot,\cdot)$ and $\widehat{h}_{2,n}(\theta_*,\cdot,\cdot)$ have the same distribution for all $n \geq 1$ and $c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha) = c_0(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha+\eta) + \eta$. *Q.E.D.*

14.3. *Proofs of Results for $n^{-1/2}$-Local Alternatives*

PROOF OF THEOREM 4: The proof of part (a) uses the following. By element-by-element mean-value expansions about θ_n and Assumptions LA1(a), LA1(b), and LA2,

$$\begin{aligned}(14.37)\quad & D_{F_n}^{-1/2}(\theta_{n,*}) E_{F_n} m(W_i, \theta_{n,*}, g) \\ & \quad = D_{F_n}^{-1/2}(\theta_n) E_{F_n} m(W_i, \theta_n, g) + \Pi_{F_n}(\theta_{n,g}, g)(\theta_{n,*} - \theta_n), \quad \text{and so} \\ & n^{1/2} D_{F_n}^{-1/2}(\theta_{n,*}) E_{F_n} m(W_i, \theta_{n,*}, g) \to h_1(g) + \Pi_0(g)\lambda,\end{aligned}$$

where $\theta_{n,g}$ may differ across rows of $\Pi_{F_n}(\theta_{n,g}, g)$, $\theta_{n,g}$ lies between $\theta_{n,*}$ and θ_n, $\theta_{n,g} \to \theta_0$, $\Pi_{F_n}(\theta_{n,g}, g) \to \Pi_0(g)$, and by definition $h_1(g) + \Pi_0(g)\lambda = \infty$ if $h_1(g) = \infty$.

Now, the proof of part (a) is the same as the proof of Theorem 2(b) with the following changes: (i) $(\theta_{n,*}, F_n)$ appears in place of (θ_c, F_c) whenever (θ_c, F_c) is used in an expression with n finite, (ii) (θ_0, F_0) appears in place of (θ_c, F_c) whenever (θ_c, F_c) is used in an asymptotic expression, (iii) $\{(\theta_{n,*}, F_n) : n \geq 1\}$ satisfies the conditions to be in SubSeq(h_2) (where $h_2 = h_{2,F_0}(\theta_0)$) by Assumptions LA1(a) and LA1(c)–(e) and because $\{W_i : i \geq 1\}$ are i.i.d. under F_n and Assumption M holds given that $(\theta_n, F_n) \in \mathcal{F}$ by Assumption LA1, (iv) (14.16) is replaced by

$$(14.38)\quad \kappa_n^{-1} \bar{D}_{F_n}^{-1/2}(\theta_{n,*}, g) D_{F_n}^{1/2}(\theta_{n,*}) h_{1,n,F_n}(\theta_{n,*}, g) \to \pi_1(g) \quad \text{as} \quad n \to \infty,$$

which holds by Assumption LA4, (14.37) (because $\kappa_n^{-1} n^{1/2} \Pi_{F_n}(\theta_{n,g}, g)(\theta_{n,*} - \theta_n) \to 0$), and $\bar{D}_{F_n}^{-1/2}(\theta_{n,*}, g)\bar{D}_{F_n}^{1/2}(\theta_n, g) \to I_k$ (using Assumption LA1(c)), (v) $\pi_1(g)$ appears in place of $h_{1,\infty,F_c}(\theta_c, g)$ in (14.17), (vi) $\varphi(\pi_1(g))$ appears in place of $h_{1,\infty,F_c}(\theta_c, g)$ in (14.18)–(14.20), where (14.18) holds for all $g \in \mathcal{G}_\varphi$ by Assumption LA5(a) and (14.19) holds for all $g \in \mathcal{G}_\varphi$, (vii) Assumption LA5(b) is used in place of Assumption GMS2(a) in two places, (viii) $(h_1 + \Pi_0 \lambda, h_2)$ and $h_1(g)$ appear in place of $h_{\infty,F_c}(\theta_c)$ and $h_{1,\infty,F_c}(\theta_c)$, respectively, in (14.23) and (14.24), and (ix) (14.23) holds using (14.37) in place of $h_{1,n,F_c}(\theta_c) \to h_{1,\infty,F_c}(\theta_c)$ and using $\nu_{n,F_n}(\theta_{n,*}, \cdot) \Rightarrow \nu_{h_2}(\cdot)$ in place of $\nu_{n,F_c}(\theta_c, \cdot) \Rightarrow \nu_{h_{2,F_c}(\theta_c)}(\cdot)$. The result $\nu_{n,F_n}(\theta_{n,*}, \cdot) \Rightarrow \nu_{h_2}(\cdot)$ holds by Lemma A1(a) because $\{(\theta_{n,*}, F_n) : n \geq 1\} \in$ SubSeq(h_2) by the argument given in (iii) above. The desired result is given by (14.24) with the changes indicated above. This completes the proof of part (a).

Part (b) holds by the same argument as for part (a) but with $\varphi_n(\theta_{n,*}, g)$ replaced by 0, which simplifies the argument considerably. Assumption LA6 is used in place of Assumption LA5(b) in the proof.

Part (c) holds by the following argument:

$$
\begin{aligned}
(14.39)\quad & \beta^{-\chi} T(h_1 + \Pi_0 \lambda_0 \beta, h_2) \\
&= \beta^{-\chi} \int S\big(\nu_{h_2}(g) + h_1(g) + \Pi_0(g)\lambda_0\beta, h_2(g) + \varepsilon I_k\big)\, dQ(g) \\
&= \int S\big(\nu_{h_2}(g)/\beta + h_1(g)/\beta + \Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k\big)\, dQ(g) \\
&\to \int S\big(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k\big)\, dQ(g) > 0
\end{aligned}
$$

as $\beta \to \infty$ a.s., where χ is as in Assumption S4, the second equality holds by Assumption S4, the convergence holds a.s. (with respect to the randomness in ν_{h_2}) by the bounded convergence theorem applied for each fixed sample path ω because $\|\nu_{h_2}(g)\|$ has bounded sample paths a.s., and using Assumption LA3′ (which guarantees that $h_{1,j}(g) < \infty$ and hence $h_{1,j}(g)/\beta \to 0$ as $\beta \to \infty$ for all $j \leq p$, for all g in a set with Q measure 1), and the inequality holds by Assumptions LA3′ and S3.

Equation (14.39) implies that $T(h_1 + \Pi_0 \lambda_0 \beta, h_2) \to \infty$ a.s. as $\beta \to \infty$. Because $T(h_1 + \Pi_0 \lambda_0 \beta, h_2) \sim J_{h,\beta\lambda_0}$ and the quantities $c(\varphi(\pi_1), h_2, 1 - \alpha)$ and $c(0_{\mathcal{G}}, h_2, 1 - \alpha)$ do not depend on β, the result of part (c) follows. *Q.E.D.*

14.4. *Proofs Concerning the Verification of Assumptions S1–S4*

PROOF OF LEMMA 1: Assumptions S1(a)–(d) and S3 hold for the functions S_1, S_2, and S_3 by Lemma 1 of Andrews and Guggenberger (2009). Assumptions S1(e) and S4 hold immediately for the functions S_1, S_2, and S_3 with $\chi = 2$ in Assumption S4.

To verify Assumption S2 for $S = S_1$, S_2, or S_3, it suffices to show that

$$\limsup_{n\to\infty} \big|S(m_n + \mu_n, \Sigma_n) - S(m_0 + \mu_n, \Sigma_0)\big| = 0 \tag{14.40}$$

for all sequences $\{\mu_n \in [0,\infty)^p \times \{0\}^v : n \geq 1\}$ and $\{(m_n, \Sigma_n) : n \geq 1\}$ such that $(m_n, \Sigma_n) \to (m_0, \Sigma_0)$, $m_0 \in R^k$, and $\Sigma_0 \in \mathcal{W}$.

For clarity of the proof, we consider a simple case first. We consider the function S_1 and suppose $\Sigma_n = \Sigma_0$. In this case, without loss of generality, we can assume that $\Sigma_0 = I_k$. Given that S_1 is additive, it suffices to consider the cases where $(p, v) = (1, 0)$ and $(0, 1)$. It is easy to see that Assumption S2 holds in the latter case because μ_n does not appear. For the case where $(p, v) = (1, 0)$, we have

$$\begin{aligned}
&\big|S_1(m_n + \mu_n, I_k) - S_1(m_0 + \mu_n, I_k)\big| \\
&\quad = \big|\big([m_n + \mu_n]_-^2 - [m_0 + \mu_n]_-^2\big)\big| \\
&\quad \leq \big|[m_n + \mu_n]_- - [m_0 + \mu_n]_-\big|\big([m_n + \mu_n]_- + [m_0 + \mu_n]_-\big) \\
&\quad \leq |m_n - m_0|\big(|m_n| + |m_0|\big) \\
&\quad = o(1)O(1),
\end{aligned} \tag{14.41}$$

where the second inequality holds because $|[a]_- - [b]_-| \leq |a - b|$ and by Assumption S1(b). This completes the verification of Assumption S2 for the simple case.

Next, we verify Assumption S2 for $S = S_2$. For any sequence $\{\mu_n \in [0,\infty)^p \times \{0\}^v : n \geq 1\}$, there exists a subsequence $\{u_n : n \geq 1\}$ of $\{n\}$ such that

$$\begin{aligned}
&\lim_{n\to\infty} \big|S_2(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) - S_2(m_0 + \mu_{u_n}, \Sigma_0)\big| \\
&\quad = \limsup_{n\to\infty} \big|S_2(m_n + \mu_n, \Sigma_n) - S_2(m_0 + \mu_n, \Sigma_0)\big|.
\end{aligned} \tag{14.42}$$

Let $\{t_{1,u_n}, t_{0,u_n} \in [0,\infty)^p \times \{0\}^v : n \geq 1\}$ be sequences such that

$$\begin{aligned}
&(m_{u_n} + \mu_{u_n} - t_{1,u_n})'\Sigma_{u_n}^{-1}(m_{u_n} + \mu_{u_n} - t_{1,u_n}) \\
&\quad \leq S_2(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) + 2^{-u_n} \quad \text{and} \\
&(m_0 + \mu_{u_n} - t_{0,u_n})'\Sigma_0^{-1}(m_0 + \mu_{u_n} - t_{0,u_n}) \leq S_2(m_0 + \mu_{u_n}, \Sigma_0) + 2^{-u_n}.
\end{aligned} \tag{14.43}$$

Then,

$$\begin{aligned}
&\lim_{n\to\infty} \big[S_2(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) - S_2(m_0 + \mu_{u_n}, \Sigma_0)\big] \\
&\quad = \lim_{n\to\infty} \big[(m_{u_n} + \mu_{u_n} - t_{1,u_n})'\Sigma_{u_n}^{-1}(m_{u_n} + \mu_{u_n} - t_{1,u_n})
\end{aligned} \tag{14.44}$$

$$
\begin{aligned}
&\quad - S_2(m_0 + \mu_{u_n}, \Sigma_0)\big] \\
&\geq \lim_{n\to\infty}\big[(m_{u_n} + \mu_{u_n} - t_{1,u_n})'\Sigma_{u_n}^{-1}(m_{u_n} + \mu_{u_n} - t_{1,u_n}) \\
&\quad - (m_0 + \mu_{u_n} - t_{1,u_n})'\Sigma_0^{-1}(m_0 + \mu_{u_n} - t_{1,u_n})\big] \\
&= \lim_{n\to\infty}\big[(m_{u_n} + \mu_{u_n} - t_{1,u_n})'\big(\Sigma_{u_n}^{-1} - \Sigma_0^{-1}\big)(m_{u_n} + \mu_{u_n} - t_{1,u_n}) \\
&\quad + (m_{u_n} - m_0)'\Sigma_0^{-1}(m_0 + m_{u_n} + 2\mu_{u_n} - 2t_{1,u_n})\big] \\
&= 0,
\end{aligned}
$$

where the last equality holds if $\mu_{u_n} - t_{1,u_n} = O(1)$ because $m_{u_n} \to m_0 < \infty$ and $\Sigma_{u_n}^{-1} \to \Sigma_0^{-1}$ as $n \to \infty$.

We now show that $\mu_{u_n} - t_{1,u_n} = O(1)$. We have

$$
\begin{aligned}
(14.45)\quad m_{u_n}'\Sigma_{u_n}^{-1}m_{u_n} &\geq S_2(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) \\
&\geq (m_{u_n} + \mu_{u_n} - t_{1,u_n})'\Sigma_{u_n}^{-1}(m_{u_n} + \mu_{u_n} - t_{1,u_n}) - 2^{-u_n}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
(14.46)\quad &\lim_{n\to\infty}(m_{u_n} + \mu_{u_n} - t_{1,u_n})'\Sigma_{u_n}^{-1}(m_{u_n} + \mu_{u_n} - t_{1,u_n}) \\
&\quad \leq \lim_{n\to\infty}\big[m_{u_n}'\Sigma_{u_n}^{-1}m_{u_n} + 2^{-u_n}\big] = m_0'\Sigma_0^{-1}m_0 < \infty,
\end{aligned}
$$

which implies that $m_{u_n} + \mu_{u_n} - t_{1,u_n} = O(1)$. The latter and $m_{u_n} \to m_0 < \infty$ give

$$
(14.47)\quad \mu_{u_n} - t_{1,u_n} = O(1).
$$

Next, by an analogous argument to (14.44) with $\geq$ and t_{1,u_n} replaced by $\leq$ and t_{0,u_n}, respectively, we obtain the following upper bound:

$$
\begin{aligned}
(14.48)\quad &\lim_{n\to\infty}\big[S(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) - S(m_0 + \mu_{u_n}, \Sigma_0)\big] \\
&= \lim_{n\to\infty}\big[S(m_{u_n} + \mu_{u_n}, \Sigma_{u_n}) \\
&\quad - (m_0 + \mu_{u_n} - t_{0,n})'\Sigma_0^{-1}(m_0 + \mu_{u_n} - t_{0,u_n})\big] \\
&\leq 0,
\end{aligned}
$$

where the inequality uses $\mu_{u_n} - t_{0,u_n} = O(1)$, which holds by an analogous argument to that given for (14.47). Equations (14.44) and (14.48) imply that the left-hand side of (14.42) equals zero, which completes the verification of Assumption S2 for S_2.

The verification of Assumption S2 for $S = S_1$, where Σ_n need not equal Σ_0, is obtained by replacing Σ_n and Σ_0 in the proof above for S_2 by $\mathrm{Diag}\{\Sigma_n\}$ and

Diag$\{\Sigma_0\}$, respectively, because $S_1(m,\Sigma)=S_2(m,\Sigma)$ when Σ is diagonal. Assumption S2 holds for the function S_3 when $(p,v)=(1,0)$ and $(0,1)$ because $S_3=S_1=S_2$ in these cases. It holds for S_3 in the general (p,v) case because it holds in these two special cases. *Q.E.D.*

15. SUPPLEMENTAL APPENDIX D

In this appendix, we provide proofs of the results stated in Supplemental Appendix B. The first subsection gives proofs for the Kolmogorov–Smirnov and approximate CvM tests and CS's. The second subsection gives proofs for results concerning $\mathcal{G}_{\text{B-spline}}$ and $\mathcal{G}_{c/d}$. The third subsection gives proofs for results concerning "asymptotic issues with the Kolmogorov–Smirnov statistic." The fourth subsection gives proofs for the subsampling results.

15.1. *Proofs of Kolmogorov–Smirnov and Approximate Cramér–von Mises Results*

PROOF OF LEMMA B1: To verify Assumption S2′ for S_1, S_2, and S_3, it suffices to show that

$$\limsup_{n\to\infty}\left|S(m_n+\mu_n,\Sigma_n)-S(m_{n,0}+\mu_n,\Sigma_{n,0})\right|=0 \tag{15.1}$$

for all sequences $\{\mu_n\in[0,\infty)^p\times\{0\}^v:n\geq 1\}$, $\{(m_n,\Sigma_n)\in\mathcal{M}\times\mathcal{W}_{\text{bd}}:n\geq 1\}$, and $\{(m_{n,0},\Sigma_{n,0})\in\mathcal{M}\times\mathcal{W}_{\text{bd}}:n\geq 1\}$ for which $(m_n,\Sigma_n)-(m_{n,0},\Sigma_{n,0})\to 0$ as $n\to\infty$.

The verification of (15.1) is an extension of the verification of (14.40) in the proof of Lemma 1. The extension consists of (i) replacing m_0 and Σ_0 by $m_{u_n,0}$ and $\Sigma_{u_n,0}$ throughout (14.42)–(14.48), (ii) making use of the fact that m_{u_n}, $m_{u_n,0}$, and $\Sigma_{u_n}^{-1}$ are bounded by the definitions of $\mathcal{M}$ and $\mathcal{W}_{\text{bd}}$, and (iii) making use of the fact that $\Sigma_{u_n}^{-1}-\Sigma_{u_n,0}^{-1}\to 0$ given that $\Sigma_{u_n}-\Sigma_{u_n,0}\to 0$ and $\Sigma_{u_n},\Sigma_{u_n,0}\in\mathcal{W}_{\text{bd}}$. *Q.E.D.*

PROOF OF THEOREM B1: When $T_n(\theta)$ is the KS statistic and when $T_n(\theta)$ is replaced by the approximate statistic $\bar{T}_{n,s_n}(\theta)$, the results of Theorem 1 hold under the assumptions of that theorem plus Assumption S2′. The proof of Theorem 1 goes through with the following changes: (i) the statistics $\tilde{T}_{a_n}$ and $\tilde{T}_{a_n,0}$ are changed from integrals with respect to Q to suprema over $g\in\mathcal{G}_n$ or weighted averages over $\{g_1,\ldots,g_{s_n}\}$ with weights $\{w_{Q,n}(\ell):\ell=1,\ldots,s_n\}$, (ii) in the proof of (12.7), (12.10) holds uniformly over $g\in\mathcal{G}$ because Assumption S2 has been strengthened to Assumption S2′, and (iii) (12.11) holds with the supremum over $g\in\mathcal{G}_n$ added or with the average over $\{g_1,\ldots,g_{s_n}\}$ added, because (12.10) holds uniformly over $g\in\mathcal{G}$ and the weights are nonnegative and sum to at most 1 by Assumption A2. This completes the proof of Theorem 1 for the KS and A-CvM test statistics.

The result of Theorem B1 is the same as that of Theorem 2(a). The proof of Theorem 2(a) follows immediately from Lemmas A2–A4. The proof of Lemma A4 uses Lemma A5. The proofs of Lemmas A2–A5 go through for the KS and A-CvM test statistics with the following minor changes: (i) in the proof of Lemma A2, $T(h)$ is replaced by $\bar{T}_{s_n}(h)$ (defined in (4.6)) and the new version of Theorem 1 for the KS and A-CvM statistics is employed, (ii) in the proof of Lemma A3, the form of the test statistic only enters through the first inequality of (12.23), which holds for the supremum and weighted average forms of the test statistic, (iii) in the proof of Lemma A4, no changes are required because the form of the test statistic only enters through Lemma A5, and (iv) in the proof of Lemma A5, $T(h)$ is replaced by $\bar{T}_{s_n}(h)$. *Q.E.D.*

PROOF OF THEOREM B2: Theorem B2 is proved by adjusting the proof Theorem 3. The proof of Theorem 3 goes through up to (14.32) with the only change being that $c(\cdot,\cdot,\cdot)$ is replaced by $c_{s_n}(\cdot,\cdot,\cdot)$ for A-CvM tests in (14.27)—in particular, the integral with respect to Q in (14.28) is not changed. Equation (14.33) needs to be replaced (see (15.2) and (15.6) below); (14.34) is established with $c(\cdot,\cdot,\cdot)$ replaced by $c_{s_n}(\cdot,\cdot,\cdot)$ for A-CvM tests; (14.35) holds, with $T_n(\theta_*)$ and $c(\cdot,\cdot,\cdot)$ replaced by $\bar{T}_{n,s_n}(\theta_*)$ and $c_{s_n}(\cdot,\cdot,\cdot)$ for A-CvM tests, using the replacements for (14.33) given in (15.2) and (15.6) below; the first equation in (14.36) holds by Lemma A5 with $c(\cdot,\cdot,\cdot)$ replaced by $c_{s_n}(\cdot,\cdot,\cdot)$ for A-CvM tests, noting that Lemma A5 is extended to KS and A-CvM critical values in the proof of Theorem B1 above; in the second equation in (14.36), "$c_0(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta) < \infty$" holds for the KS critical value because $c_0(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta)$ does not depend on n and the KS test statistic $T(0_{\mathcal{G}}, h_{2,0})$ is finite a.s. since the sample paths of $\nu_{h_{2,0}}(\cdot)$ and $h_{2,0}(\cdot)$ are bounded a.s.; and in the second equation in (14.36), "$\sup_{n\geq 1} c_{0,s_n}(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta) < \infty$" holds for an A-CvM critical value because $c_{0,s_n}(0_{\mathcal{G}}, h_{2,0}, 1-\alpha+\eta+\delta)$ is less than or equal to the corresponding quantile based on the KS statistic, which does not depend on n and is finite a.s.

For the KS test, we replace (14.33) with the following:

$$
\begin{aligned}
(15.2)\quad & \big(n^{1/2}\beta(g_0)\big)^{-\chi} T_n(\theta_*)\cdot Q\big(\mathcal{B}_{\rho_X}(g_0,\tau_2)\big) \\
&= \big(n^{1/2}\beta(g_0)\big)^{-\chi} \sup_{g\in\mathcal{G}_n} S\big(\nu_{n,F_0}(\theta_*,g)+h_{1,n,F_0}(\theta_*,g), \bar{h}_{2,n,F_0}(\theta_*,g)\big) \\
&\quad \times Q\big(\mathcal{B}_{\rho_X}(g_0,\tau_2)\big) \\
&\geq \big(n^{1/2}\beta(g_0)\big)^{-\chi} \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} 1(g\in\mathcal{G}_n) \\
&\quad \times S\big(\nu_{n,F_0}(\theta_*,g)+h_{1,n,F_0}(\theta_*,g), \bar{h}_{2,n,F_0}(\theta_*,g)\big)\, dQ(g)
\end{aligned}
$$

$$= \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} 1(g \in \mathcal{G}_n) S\big(\big(n^{1/2}\beta(g_0)\big)^{-1}\nu_{n,F_0}(\theta_*, g)$$
$$+ m^*(g)/\beta(g_0), \bar{h}_{2,n,F_0}(\theta_*, g)\big)\, dQ(g)$$
$$\to_p \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\big(m^*(g)/\beta(g_0), h_{2,0}(g) + \varepsilon I_k\big)\, dQ(g) > 0,$$

where χ is as in Assumption S4, $m^*(g) = (m_1^*(g), \dots, m_k^*(g))'$, $m_j^*(g)$ is defined in (6.2) for $j \le k$, $h_{2,0} = h_{2,F_0}(\theta_*)$, and the convergence uses the argument given in the paragraph following (14.33) as well as $1(g \in \mathcal{G}_n) \to 1(g \in \mathcal{G}) = 1$ as $n \to \infty$ by Assumption KS.

For A-CvM tests, we replace (14.33) with the following results:

$$(15.3)\quad \big(n^{1/2}\beta(g_0)\big)^{-\chi}\bar{T}_{n,s_n}(\theta_*)$$
$$= \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S\big(\big(n^{1/2}\beta(g_0)\big)^{-1}\nu_{n,F_0}(\theta_*, g_\ell)$$
$$+ m^*(g_\ell)/\beta(g_0), \bar{h}_{2,n,F_0}(\theta_*, g_\ell)\big),$$

using Assumption S4. We have

$$(15.4)\quad \sup_{g\in\mathcal{G}}\big|m_j^*(g)\big| \le \big(E_{F_0} m_j^2(W_i, \theta_*)/\sigma_{F_0,j}^2(\theta_*)\big)^{1/2}\big(E_{F_0} G^2(X_i)\big)^{1/2} < \infty,$$

for $j = 1, \dots, k$, using the definition of $m^*(g)$, Assumption FA (which imposes Assumption M in part FA(e)), and the Cauchy–Schwarz inequality. Next, we have

$$(15.5)\quad \sup_{g\in\mathcal{G}}\big|S\big(\big(n^{1/2}\beta(g_0)\big)^{-1}\nu_{n,F_0}(\theta_*, g) + m^*(g)/\beta(g_0), \bar{h}_{2,n,F_0}(\theta_*, g)\big)$$
$$- S\big(m^*(g)/\beta(g_0), h_{2,0}(g) + \varepsilon I_k\big)\big| = o_p(1)$$

under F_0, using the uniform continuity of S over a compact set, which holds by Assumption S1(d), where attention can be restricted to a compact set by (i) equation (15.4), (ii) $\sup_{g\in\mathcal{G}} \|n^{-1/2}\nu_{n,F_0}(\theta_*, g)\| = o_p(1)$ by Lemma A1(a), and (iii) $\sup_{g\in\mathcal{G}} \|\bar{h}_{2,n,F_0}(\theta_*) - h_{2,0} - \varepsilon I_k\| = o_p(1)$ using Lemma A1(b) and the definition of $\bar{h}_{2,n,F_0}(\theta_*)$ in (5.2), and Lemma A1 applies for the reasons given in the paragraph following (14.33).

Equations (15.3) and (15.5) yield

$$(15.6)\quad \big(n^{1/2}\beta(g_0)\big)^{-\chi}\bar{T}_{n,s_n}(\theta_*) + o_p(1)$$
$$= \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S\big(m^*(g_\ell)/\beta(g_0), h_{2,0}(g_\ell)\big)$$

$$\to \int S\big(m^*(g)/\beta(g_0), h_{2,0}(g)\big)\, dQ(g)$$

$$\geq \int_{\mathcal{B}_{\rho_X}(g_0,\tau_2)} S\big(m^*(g)/\beta(g_0), h_{2,0}(g)\big)\, dQ(g) > 0,$$

where the convergence holds for fixed $\{g_1, g_2, \ldots\}$ by Assumptions A1, A2, and S4, the first inequality holds by Assumption S1(c), and the second inequality holds by (14.28). This completes the proof. *Q.E.D.*

PROOF OF THEOREM B3: Part (a) follows from part (b) because

$$c_{s_n}\big(\varphi_n(\theta_{n,*}), \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha\big) \leq c_{s_n}\big(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha\big), \tag{15.7}$$

which holds because $\varphi_n(\theta_*, g) \geq 0_k\ \forall g \in \mathcal{G}$ by Assumption GMS1(a), $c(h_1, \widehat{h}_{2,n}(\theta_*), 1-\alpha)$ is non-increasing in the first p elements of h_1 by Assumption S1(b), and the last v elements of $\varphi_n(\theta_*, g)$ equal zero.

Now, we prove part (b). When $T_n(\theta)$ is replaced by the A-CvM statistic $\bar{T}_{n,s_n}(\theta_{n,*})$, the results of Theorem 1 hold under Assumptions M, S1, and S2′ with (θ, F) replaced by $(\theta_{n,*}, F_n)$, $\sup_{(\theta,F)\in\mathcal{F}:h_{2,F}(\theta)\in\mathcal{H}_{2,\text{cpt}}}$ deleted, $T_n(\theta)$, $T(h_{n,F}(\theta))$, and $x_{h_{n,F}(\theta)}$ replaced by $\bar{T}_{n,s_n}(\theta_{n,*})$, $\bar{T}_{s_n}(h_{n,F_n}(\theta_{n,*}))$ (defined in (4.6)), and $x_{h_{n,F_n}(\theta_{n,*})}$, respectively, where $x_{h_{n,F_n}(\theta_{n,*})} \in R$ is a constant that may depend on $(\theta_{n,*}, F_n)$ and n through $h_{n,F_n}(\theta_{n,*})$. The adjustments needed to the proof of Theorem 1 are quite similar to those stated at the beginning of the proof of Theorem B1. In addition, the proof uses the fact that $\{(\theta_{n,*}, F_n) : n \geq 1\}$ satisfies the conditions to be in SubSeq(h_2) (where $h_2 = h_{2,F_0}(\theta_0)$) by Assumptions LA1(a) and LA1(c)–(e) and because $\{W_i : i \geq 1\}$ are i.i.d. under F_n and Assumption M holds given that $(\theta_n, F_n) \in \mathcal{F}$ by Assumption LA1. Because $\{(\theta_{n,*}, F_n) : n \geq 1\} \in$ SubSeq(h_2), Lemma A1 applies, which is used in (12.3). Also, $(h_{1,n,F}(\theta), h_{2,F}(\theta))$ is changed to $(h_{1,n,F_n}(\theta_{n,*}), h_{2,F_n}(\theta_{n,*}))$ throughout the proof of Theorem 1.

Next, using the mean-value expansion in (14.37) and the definition $h_{1,n,F}(\theta, g) = n^{1/2} D_F^{-1/2}(\theta) E_F m(W_i, \theta, g)$, we have

$$\begin{aligned}
&\sup_{g\in\mathcal{G}} \big\| h_{1,n,F_n}(\theta_{n,*}, g) - h_{1,n,F_n}(\theta_n, g) - \Pi_0(g)\lambda \big\| \\
&\quad = \sup_{g\in\mathcal{G}} \big\| \Pi_{F_n}(\theta_{n,g}, g) n^{1/2}(\theta_{n,*} - \theta_n) - \Pi_0(g)\lambda \big\| \\
&\quad \leq \sup_{g\in\mathcal{G}} \sup_{\theta\in\Theta:\|\theta-\theta_0\|\leq\delta_n} \big\| \Pi_{F_n}(\theta, g)\lambda\big(1 + o(1)\big) - \Pi_0(g)\lambda \big\| \\
&\quad \to 0,
\end{aligned} \tag{15.8}$$

where $\theta_{n,g}$ may differ across rows of $\Pi_{F_n}(\theta_{n,g}, g)$, $\theta_{n,g}$ lies between $\theta_{n,*}$ and θ_n, $\delta_n = \|\theta_{n,*} - \theta_n\| + \|\theta_n - \theta_0\| \to 0$, the inequality holds using Assumption LA1(a), and the convergence to zero uses Assumption LA2′(b). (Note that the $(1+o(1))$ term in (15.8) requires the condition in Assumption LA2′(b) that $\sup_{g\in\mathcal{G}}\|\Pi_0(g)\lambda\|<\infty$).

Equation (15.8) and Assumption LA2′(a) give: for all $B<\infty$,

$$\sup_{g\in\mathcal{G}:h_1(g)\leq B}\big\|h_{1,n,F_n}(\theta_{n,*},g)-h_1(g)-\Pi_0(g)\lambda\big\|\to 0. \tag{15.9}$$

By Assumption LA1(c), $d(h_{2,F_n}(\theta_{n,*}),h_{2,F_0}(\theta_0))\to 0$. This implies that $\nu_{h_{2,F_n}(\theta_{n,*})}(\cdot)\Rightarrow\nu_{h_2}(\cdot)$, where $h_2=h_{2,F_0}(\theta_0)$. As in previous proofs, by the almost sure representation theorem, there exist a probability space and random quantities $\tilde{\nu}_n(\cdot)$ and $\tilde{\nu}(\cdot)$ defined on it with the same distributions as $\nu_{h_{2,F_n}(\theta_{n,*})}(\cdot)$ and $\nu_{h_2}(\cdot)$, respectively, for $n\geq 1$, such that $\sup_{g\in\mathcal{G}}\|\tilde{\nu}_n(g)-\tilde{\nu}(g)\|\to 0$ a.s. Hence, $\bar{T}_{s_n}(h_{n,F_n}(\theta_{n,*}))$ and $\bar{\bar{T}}_{s_n}(h_{n,F_n}(\theta_{n,*}))$ have the same distribution, where the latter is defined to be

$$\begin{aligned}&\bar{\bar{T}}_{s_n}\big(h_{n,F_n}(\theta_{n,*})\big)\\&\quad=\sum_{\ell=1}^{s_n}w_{Q,n}(\ell)S\big(\tilde{\nu}_n(g_\ell)+h_{1,n,F_n}(\theta_{n,*},g_\ell),h_{2,F_n}(\theta_{n,*},g_\ell)+\varepsilon I_k\big).\end{aligned}\tag{15.10}$$

For all $\beta>0$, $B<\infty$, and $\lambda=\lambda_0\beta$, we have

$$\begin{aligned}A_{1,n}(\beta,B)=\sup_{g\in\mathcal{G}:h_1(g)\leq B}&\big|S\big(\tilde{\nu}_n(g)/\beta\\&+h_{1,n,F_n}(\theta_{n,*},g)/\beta,h_{2,F_n}(\theta_{n,*},g)+\varepsilon I_k\big)\\&-S\big(\tilde{\nu}(g)/\beta+h_1(g)/\beta+\Pi_0(g)\lambda_0,h_2(g)+\varepsilon I_k\big)\big|\\&\to 0\quad\text{as}\quad n\to\infty\quad\text{a.s.}\end{aligned}\tag{15.11}$$

using Assumption S2′, (15.9), $\sup_{g\in\mathcal{G}}\|\tilde{\nu}_n(g)-\tilde{\nu}(g)\|\to 0$ a.s., $\sup_{g\in\mathcal{G}}\|\tilde{\nu}(g)\|<\infty$ a.s., and $d(h_{2,F_n}(\theta_{n,*}),h_2)\to 0$, where $h_2=h_{2,F_0}(\theta_0)$.

In addition, for all $B<\infty$, we have

$$\begin{aligned}A_2(\beta,B)=\sup_{g\in\mathcal{G}:h_1(g)\leq B}&\big|S\big(\tilde{\nu}(g)/\beta+h_1(g)/\beta+\Pi_0(g)\lambda_0,h_2(g)+\varepsilon I_k\big)\\&-S\big(\Pi_0(g)\lambda_0,h_2(g)+\varepsilon I_k\big)\big|\\&\to 0\quad\text{as}\quad\beta\to\infty\quad\text{a.s.}\end{aligned}\tag{15.12}$$

We use (15.11) and (15.12) to obtain: for all constants $B_c^*<\infty$ as in Assumption A3,

$$\begin{aligned}&\beta^{-\chi}\bar{\bar{T}}_{s_n}\big(h_{n,F_n}(\theta_{n,*})\big)\\&\quad\geq\sum_{\ell=1}^{s_n}w_{Q,n}(\ell)1\big(h_1(g_\ell)\leq B_c^*\big)\end{aligned}\tag{15.13}$$

$$
\begin{aligned}
&\times S\big(\tilde{\nu}_n(g_\ell)/\beta + h_{1,n,F_n}(\theta_{n,*}, g_\ell)/\beta, h_{2,F_n}(\theta_{n,*}, g_\ell) + \varepsilon I_k\big) \\
\geq &\sum_{\ell=1}^{s_n} w_{Q,n}(\ell) 1\big(h_1(g_\ell) \leq B_c^*\big) S\big(\Pi_0(g_\ell)\lambda_0, h_2(g_\ell) + \varepsilon I_k\big) \\
&- A_{1,n}\big(\beta, B_c^*\big) - A_2\big(\beta, B_c^*\big) \\
\to_{n\to\infty} &\text{ a.s. } \int 1\big(h_1(g) \leq B_c^*\big) S\big(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k\big)\, dQ(g) \\
&\qquad - A_2\big(\beta, B_c^*\big) \\
\to_{\beta\to\infty} &\text{ a.s. } \int 1\big(h_1(g) \leq B_c^*\big) S\big(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k\big)\, dQ(g),
\end{aligned}
$$

where the first inequality uses Assumptions S1(c) and S4, the second inequality holds by the definitions of $A_{1,n}(\beta, B_c^*)$ and $A_2(\beta, B_c^*)$, the first convergence result holds by (15.11) and Assumption A3, and the second convergence result holds by (15.12).

Let $c_{\sup,0}(0_\mathcal{G}, h_2^*, 1-\alpha)$ denote the $1-\alpha$ quantile of $T_{\sup}(0_\mathcal{G}, h_2^*) = \sup_{g\in\mathcal{G}} S(\nu_{h_2}(g), h_2^*(g) + \varepsilon I_k)$, where h_2^* is some $k \times k$-matrix-valued covariance kernel on $\mathcal{G} \times \mathcal{G}$. Let $0_{\mathcal{G}\times\mathcal{G}}$ denote the $k \times k$-matrix-valued covariance kernel on $\mathcal{G} \times \mathcal{G}$ that equals the $k \times k$ zero matrix for all $(g, g^*) \in \mathcal{G} \times \mathcal{G}$. The A-PA critical value satisfies

$$
\begin{aligned}
(15.14)\qquad c_{s_n}\big(0_\mathcal{G}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha\big) &\leq c_{\sup,0}\big(0_\mathcal{G}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha+\eta\big) + \eta \\
&\leq c_{\sup,0}(0_\mathcal{G}, 0_{\mathcal{G}\times\mathcal{G}}, 1-\alpha+\eta) + \eta \\
&< \infty,
\end{aligned}
$$

where the first inequality holds because a weighted average over $\{g_1, \ldots, g_{s_n}\}$ with nonnegative weights that sum to 1 or less (by Assumption A2) is less than or equal to the corresponding supremum over $g \in \mathcal{G}$, which implies that $\bar{T}_{s_n}(0_\mathcal{G}, h_2^*) \leq T_{\sup}(0_\mathcal{G}, h_2^*)\ \forall h_2^*$, the second inequality holds because $S(\nu_{h_2}(g), h_2^*(g) + \varepsilon I_k) \leq S(\nu_{h_2}(g), \varepsilon I_k)\ \forall g \in \mathcal{G}$, for all covariance kernels h_2^* by Assumption S1(e), which implies that $T_{\sup}(0_\mathcal{G}, h_2^*) \leq T_{\sup}(0_\mathcal{G}, 0_{\mathcal{G}\times\mathcal{G}})\ \forall h_2^*$, and the last inequality holds because $\sup_{g\in\mathcal{G}} S(\nu_{h_2}(g), \varepsilon I_k) < \infty$ a.s., which holds by Assumption S2′ and $\sup_{g\in\mathcal{G}} \|\nu_{h_2}(g)\| < \infty$ a.s.

We now have: for all B_c^* as in Assumption A3,

$$
\begin{aligned}
(15.15)\qquad &\lim_{\beta\to\infty} \liminf_{n\to\infty} P_{F_n}\big(\bar{T}_{s_n}\big(h_{n,F_n}(\theta_{n,*})\big) > c_{s_n}\big(0_\mathcal{G}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha\big)\big) \\
&\geq \lim_{\beta\to\infty} \liminf_{n\to\infty} P\big(\beta^{-\chi} \bar{\bar{T}}_{s_n}\big(h_{n,F_n}(\theta_{n,*})\big) \\
&\qquad > \beta^{-\chi} c(0_\mathcal{G}, 0_{\mathcal{G}\times\mathcal{G}}, 1-\alpha+\eta) + \beta^{-\chi}\eta\big)
\end{aligned}
$$

$$\geq \lim_{\beta\to\infty} P\bigg(\int 1\big(h_1(g) \leq B_c^*\big) S\big(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k\big)\, dQ(g)$$

$$- A_2\big(\beta, B_c^*\big) > \beta^{-\chi} c(0_{\mathcal{G}}, h_2, 1-\alpha+\eta) + \beta^{-\chi}\eta\bigg)$$

$$= 1\bigg(\int 1\big(h_1(g) \leq B_c^*\big) S\big(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k\big)\, dQ(g) > 0\bigg),$$

where the first inequality holds by (15.14) and the equality in distribution of $\bar{\bar{T}}_{s_n}(h_{n,F_n}(\theta_{n,*}))$ and $\bar{T}_{s_n}(h_{n,F_n}(\theta_{n,*}))$, the second inequality holds by (i) the first two inequalities in (15.13), (ii) the first convergence result in (15.13), and (iii) the bounded convergence theorem, and the equality holds by the second convergence result of (15.13) and the bounded convergence theorem.

The left-hand side (l.h.s.) in (15.15) does not depend on B_c^*. Hence, the l.h.s. is greater than or equal to the limit as $c \to \infty$ of the right-hand side, which equals

$$1\bigg(\int 1\big(h_1(g) \leq \infty\big) S\big(\Pi_0(g)\lambda_0, h_2(g) + \varepsilon I_k\big)\, dQ(g) > 0\bigg) = 1 \tag{15.16}$$

by the monotone convergence theorem and the assumption that $B_c^* \to \infty$ as $c \to \infty$, where the equality holds by Assumptions LA3′ and S3.

Lastly, we prove part (c) regarding KS tests and CS's. The proof is essentially the same as that for parts (a) and (b) with $\bar{T}_{n,s_n}(\theta_{n,*})$, $c_{s_n}(\cdot,\cdot,\cdot)$, $\sum_{\ell=1}^{s_n} w_{Q,n}(\ell)\cdots$, and $\int\cdots dQ(g)$ replaced by the KS quantities $T_n(\theta_{n,*})$, $c(\cdot,\cdot,\cdot)$, $\sup_{g\in\mathcal{G}}$, and $\sup_{g\in\mathcal{G}}\cdots$, respectively (or with $\mathcal{G}_n$ in place of $\mathcal{G}$). *Q.E.D.*

15.2. *Proof of Lemma B2 Regarding* $\mathcal{G}_{\text{B-spline}}$, $\mathcal{G}_{\text{box,dd}}$, *and* $\mathcal{G}_{c/d}$

PROOF OF LEMMA B2: First we verify Assumption CI for $\mathcal{G} = \mathcal{G}_{\text{B-spline}}$. Let $m_{j,F}(\theta, x) = E_F(m_j(W_i, \theta)|X_i = x)$. Write

$$\mathcal{X}_F(\theta) = \left(\bigcup_{j=1}^{p} \{x \in R^{d_x} : m_{j,F}(\theta, x) < 0\}\right) \cup \left(\bigcup_{j=p+1}^{k} \{x \in R^{d_x} : m_{j,F}(\theta, x) \neq 0\}\right). \tag{15.17}$$

If $P_F(X_i \in \mathcal{X}_F(\theta)) > 0$, then the probability that X_i lies in one of the k sets in (15.17) is positive. Suppose (without loss of generality) that $P_F(X_i \in$

$\{x : m_{1,F}(\theta, x) < 0\}) > 0$. The set $\{x : m_{1,F}(\theta, x) < 0\}$ can be written as the union of disjoint nondegenerate hypercubes in $\mathcal{C}_{\text{B-spline}}$ (i.e., hypercubes with positive Lebesgue volumes) because continuity of $m_{1,F}(\theta, x)$ implies that if $m_{1,F}(\theta, x) < 0$, then $m_{1,F}(\theta, y) < 0$ for all y in some hypercube that includes x. The number of such hypercubes is countable (because otherwise their union would have infinite volume). One of these hypercubes, call it H, must have positive X_i probability. (Otherwise, the union of these hypercubes would have X_i probability zero.)

In sum, we have $H \in \mathcal{C}_{\text{B-spline}}$, $P_F(X_i \in H) > 0$, and $m_{1,F}(\theta, x) < 0$ for all $x \in H$. In addition, the B-spline whose support is H is positive on the interior of H. Thus, if $P_F(X_i \in \text{int}(H)) > 0$, we have $E_F m_1(W_i, \theta) B_H(X_i) < 0$, which establishes Assumption CI.

On the other hand, if $P_F(X_i \in \text{int}(H)) = 0$, then we must have $P_F(X_i \in H \setminus \text{int}(H)) > 0$. Because $m_{1,F}(\theta, x)$ is a continuous function of x, there exists a finite number of hypercubes in $\mathcal{C}_{\text{B-spline}}$ whose interiors have union that includes $H \setminus \text{int}(H)$ and for which $m_{1,F}(\theta, x) < 0$ for all x in each hypercube. One of these hypercubes, say H_1, must have interior with positive probability because $P_F(X_i \in H \setminus \text{int}(H)) > 0$. In sum, $H_1 \in \mathcal{C}_{\text{B-spline}}$, $P_F(X_i \in \text{int}(H_1)) > 0$, $m_{1,F}(\theta, x) < 0$ for all $x \in H_1$, and the B-spline $B_{H_1}(x)$ is positive for $x \in \text{int}(H_1)$. Hence, $E_F m_1(W_i, \theta) B_{H_1}(X_i) < 0$, which establishes Assumption CI.

Now we establish Assumption CI for $\mathcal{G}_{\text{box,dd}}$. The fact that Assumption CI holds for $\mathcal{G} = \mathcal{G}_{\text{box}}$ for all $\bar{r} \in (0, \infty]$ by Lemma 3 implies that Assumption CI holds for $\mathcal{G} = \mathcal{G}_{\text{box,dd}}$ for all $\bar{r} \in (0, \infty]$. The reason is as follows. Let $\mathcal{G}_{\text{box}}(\bar{r})$ and $\mathcal{G}_{\text{box,dd}}(\bar{r})$ denote $\mathcal{G}_{\text{box}}$ and $\mathcal{G}_{\text{box,dd}}$, respectively, when $\bar{r}$ is the upper bound on r_u or $r_{1,u}$ and $r_{2,u}$. For any box $C_{x_0,r} \in \mathcal{G}_{\text{box}}(\bar{r})$, if $C_{x_0,r}$ captures some deviation from the model, that is, $E_F m_j(W_i, \theta) 1(X_i \in C_{x_0,r}) < 0$ for some $j = 1, \ldots, p$ or $E_F m_j(W_i, \theta) 1(X_i \in C_{x_0,r}) \neq 0$ for some $j = p+1, \ldots, k$, then (i) $C_{x_0,r} \cap \text{Supp}_{F_{X,0}}(X_i) \neq \emptyset$ and (ii) $C_{x_0+\eta,r+\eta}$ captures the same deviation for $\eta > 0$ sufficiently small. Result (ii) holds because $\lim_{\eta \downarrow 0} E_F m_j(W_i, \theta) 1(X_i \in C_{x_0+\eta,r+\eta}) = E_F m_j(W_i, \theta) 1(X_i \in C_{x_0,r})$. The latter holds by the bounded convergence theorem because $(C_{x_0+\eta,r+\eta} - C_{x_0,r}) \downarrow \emptyset$ as $\eta \downarrow 0$, and hence $m_j(w, \theta) 1(x \in C_{x_0+\eta,r+\eta}) \to m_j(w, \theta) 1(x \in C_{x_0,r})$ as $\eta \downarrow 0$ for every w, and $E_F |m_j(W_i, \theta) 1(X_i \in C_{x_0+\eta,r+\eta})| \leq E_F |m_j(W_i, \theta)| < \infty$. By (i) and $\eta \in (0, \bar{r}/2]$, $C_{x_0+\eta,r+\eta}$ can be written as a box, C_{x,r_1,r_2} in $\mathcal{G}_{\text{box,dd}}(3\bar{r})$ by picking a point $x \in C_{x_0,r} \cap \text{Supp}_{F_{X,0}}(X_i)$, which is necessarily in the interior of $C_{x_0+\eta,r+\eta}$, and letting $r_1 = x - x_0 + r$ and $r_2 = x_0 + r - x + 2\eta$. We have $|x - x_0| \leq \bar{r}$, $r_1 \leq 2\bar{r}$, and $r_2 \leq 3\bar{r}$. Because $C_{x,r_1,r_2} = C_{x_0+\eta,r+\eta}$ and $C_{x_0+\eta,r+\eta}$ captures a deviation from the model, C_{x,r_1,r_2} does as well, and the proof is complete.

Note that in the preceding argument, it is necessary to expand $C_{x_0,r}$ to $C_{x_0+\eta,r+\eta}$ because $C_{x_0,r}$ is not necessarily in $\mathcal{G}_{\text{box,dd}}(3\bar{r})$ if the only elements of $C_{x_0,r} \cap \text{Supp}_{F_{X,0}}(X_i)$ are on the boundary of $C_{x_0,r}$. Also, note that the argument above does not go through if one uses symmetric side lengths (i.e., $r_{1,u} = r_{2,u}$) in the definition of $\mathcal{G}_{\text{box,dd}}$.

Next, we verify Assumption CI for $\mathcal{G} = \mathcal{G}_{c/d}$. We write

$$\mathcal{X}_F(\theta) = \bigcup_{d\in D} \mathcal{X}_{1,F}(\theta, d), \quad \text{where} \tag{15.18}$$
$$\begin{aligned}\mathcal{X}_{1,F}(\theta, d) = \big\{x_1 \in R^{d_{x,1}} : {} & E_F\big(m_j(W_i, \theta)|X_{1,i} = x_1, X_{2,i} = d\big) < 0 \\ & \text{for some } j \le p \text{ or} \\ & E_F\big(m_j(W_i, \theta)|X_{1,i} = x_1, X_{2,i} = d\big) \neq 0 \\ & \text{for some } j = p+1, \ldots, k\big\},\end{aligned}$$

for $d \in D$. We have

$$\begin{aligned}P_F\big(X_i \in \chi_F(\theta)\big) &= P_F\Big(\big(X'_{1,i}, X'_{2,i}\big)' \in \bigcup_{d\in D} \chi_{1,F}(\theta, d)\Big) \\ &= \sum_{d\in D} P_F\big(X_{1,i} \in \chi_{1,F}(\theta, d)|X_{2,i} = d\big) P_F(X_{2,i} = d).\end{aligned} \tag{15.19}$$

If $P_F(X_i \in \chi_F(\theta)) > 0$, then there exists some $d^* \in D$ such that $P_F(X_{2,i} = d^*) > 0$ and

$$P_F\big(X_{1,i} \in \chi_{1,F}\big(\theta, d^*\big)|X_{2,i} = d^*\big) > 0. \tag{15.20}$$

Given the inequality in (15.20), we use the same argument to verify Assumption CI as given for $\mathcal{G}_{\text{c-cube}}$, $\mathcal{G}_{\text{box}}$, $\mathcal{G}_{\text{B-spline}}$, or $\mathcal{G}_{\text{box,dd}}$ with d_x replaced by $d_{x,1}$, but with $E_F(\cdot)$ replaced by $E_F(\cdot|X_{2,i} = d^*)$ throughout, and using the fact that $\{g : g = g_1 1_{\{d^*\}}, g_1 \in \mathcal{G}_1\} \subset \mathcal{G}_{c/d}$ for $\mathcal{G}_1 = \mathcal{G}_{\text{c-cube}}, \mathcal{G}_{\text{box}}, \mathcal{G}_{\text{B-spline}}$, or $\mathcal{G}_{\text{box,dd}}$.

Next, we verify Assumption M. Assumptions M(a) and M(b) hold for $\mathcal{G}_{\text{B-spline}}$ by taking $G(x) = 2/3\ \forall x$ and $\delta_1 = 4/\delta + 3$. Assumption M(c) holds for $\mathcal{G}_{\text{B-spline}}$ because each element of $\mathcal{G}_{\text{B-spline}}$ can be written as the sum of four functions, each of which is the product of an indicator function of a box and a polynomial of order 4. Manageability of polynomials and indicator functions of boxes hold because they have finite pseudo-dimension (as defined in Pollard (1990, Sec. 4)). Manageability of finite linear combinations of these functions holds by the stability properties of cover numbers under addition and pointwise multiplication; see Pollard (1990, Sec. 5).

Assumption M holds for $\mathcal{G}_{\text{box,dd}}$ because it holds for $\mathcal{G}_{\text{box}}$ by Lemma 3 and $\mathcal{G}_{\text{box,dd}} \subset \mathcal{G}_{\text{box}}$.

The verification of Assumption M for $\mathcal{G} = \mathcal{G}_{c/d}$ is the same as in the proof of Lemma 3 when $\mathcal{G}_1$ is $\mathcal{G}_{\text{c-cube}}$, $\mathcal{G}_{\text{box}}$, or $\mathcal{G}_{\text{box,dd}}$ because $\mathcal{C}_{\text{box}} \times \{\{d\} : d \in D\}$ is a Vapnik–Cervonenkis class of sets. The verification of Assumption M for $\mathcal{G} = \mathcal{G}_{c/d}$ when $\mathcal{G}_1$ is $\mathcal{G}_{\text{B-spline}}$ is essentially the same as the proof above for $\mathcal{G}_{\text{B-spline}}$. The functions in $\mathcal{G}_{c/d}$ in this case still can be written as the sum of four functions, each of which is the product of an indicator function of a box—in this case, the box is of the form $B \times \{d\}$, where B is a box in $R^{d_{x,1}}$ and $d \in D$—and a polynomial of order 4.

Assumption FA(e) holds for $\mathcal{G}_{\text{B-spline}}$, $\mathcal{G}_{\text{box,dd}}$, and $\mathcal{G}_{c/d}$ by the same arguments as given above for Assumption M.

This completes the proofs of parts (a)–(d) of the lemma.

Part (e) of the lemma holds, that is, $\text{Supp}(Q_c) = \mathcal{G}_{\text{B-spline}}$, because $\mathcal{G}_{\text{B-spline}}$ is countable and Q_c has a probability mass function that is positive at each element in $\mathcal{G}_{\text{B-spline}}$.

Now, we prove part (f) using a similar argument to that for part (b) of Lemma 4. Consider $g = g_{x,r_1,r_2} \in \mathcal{G}_{\text{box,dd}}$, where $g_{x,r_1,r_2}(y) = 1(y \in C_{x,r_1,r_2}) \cdot 1_k$ and $(x, r_1, r_2) \in \text{Supp}(X_i) \times (\bigtimes_{u=1}^{d_x}(0, \sigma_{X,u}\bar{r}))^2$. Let $\delta > 0$ be given. Let $\eta_0 = (\eta_{0,1}, \ldots, \eta_{0,d_x})'$ and likewise for η_1 and η_2. Define

$$G_{g,\bar{\eta}} = \{g_{x+\eta_0,r_1-\eta_1,r_2+\eta_2} : -\bar{\eta} \le \eta_{0,u} \le \bar{\eta},\ \bar{\eta} \le \eta_{1,u},\ \eta_{2,u} \le 2\bar{\eta}\ \forall u \le d_x\}. \tag{15.21}$$

By the same sort of argument as for (14.26), for $g^* = g_{x+\eta_0,r_1-\eta_1,r_2+\eta_2} \in G_{g,\bar{\eta}}$, we have

$$\begin{aligned} \rho_X^2(g, g^*) &= E_{F_{X,0}}\big[1(X_i \in C_{x,r_1,r_2}) - 1(X_i \in C_{x+\eta_0,r_1-\eta_1,r_2+\eta_2})\big]^2 \\ &\le \sum_{u=1}^{d_x}\big[P_{F_{X,0}}\big(X_{i,u} \in \big(x_u - r_{1,u}, x_u + \eta_{0,u} - (r_{1,u} - \eta_{1,u})\big]\big) \\ &\quad + P_{F_{X,0}}\big(X_{i,u} \in (x_u + r_{2,u}, x_u + \eta_{0,u} + r_{2,u} + \eta_{2,u}]\big)\big] \\ &\le \sum_{u=1}^{d_x}\big[F_{X_u,0}(x_u - r_{1,u} + 3\bar{\eta}) - F_{X_u,0}(x_u - r_{1,u})\big] \\ &\quad + \sum_{u=1}^{d_x}\big[F_{X_u,0}(x_u + r_{2,u} + 3\bar{\eta}) - F_{X_u,0}(x_u + r_{2,u})\big], \end{aligned} \tag{15.22}$$

where $F_{X_u,0}(\cdot)$ denotes the distribution function of $X_{i,u}$ and the first inequality holds because $\eta_{0,u} + \eta_{1,u} \ge 0$ and $\eta_{0,u} + \eta_{2,u} \ge 0$. Because distribution functions are right-continuous, the r.h.s. of (15.22) converges to zero as $\bar{\eta} \downarrow 0$. Thus, $\rho_X^2(g, g^*)$ converges to zero uniformly over $G_{g,\bar{\eta}}$ as $\bar{\eta} \downarrow 0$ and there exists an $\bar{\eta} > 0$ sufficiently small that $G_{g,\bar{\eta}} \subset \mathcal{B}_{\rho_X}(g, \delta)$.

Next, we have $Q_c(G_{g,\bar{\eta}})$ equals

$$\begin{aligned} &Q^*_{F_{X,0}}\Bigg(\bigtimes_{u=1}^{d_x}[x_u - \bar{\eta}, x_u + \bar{\eta}] \bigtimes_{u=1}^{d_x}[r_{1,u} - 2\bar{\eta}, r_{1,u} - \bar{\eta}] \\ &\quad \times \bigtimes_{u=1}^{d_x}[r_{2,u} + \bar{\eta}, r_{2,u} + 2\bar{\eta}]\Bigg) > 0, \end{aligned} \tag{15.23}$$

where $Q^*_{F_{X,0}} = F_{X,0} \times \mathrm{Unif}((\times_{u=1}^{d_x}(0, \sigma_{X,u}\bar{r}))^2)$ and the inequality holds because $x \in \mathrm{Supp}(X_i)$ and $\bar{\eta} > 0$. This completes the proof of part (f).

Lastly, we prove part (g). By parts (e) and (f) and parts (a) and (b) of Lemma 4, we have $\mathcal{G}_1 \subset \mathrm{Supp}(Q_1)$. Because $\mathrm{Supp}(Q_D) = D$ and $Q_e = Q_1 \times Q_D$, we have $\mathcal{G}_{c/d} \subset \mathrm{Supp}(Q_e)$. *Q.E.D.*

15.3. *Proofs of Theorems B4 and B5 Regarding Uniformity Issues*

PROOF OF THEOREM B4: Part (a) holds by an empirical process central limit theorem because the intervals $\{(a, b] : 0 \le a < b \le 1\}$ form a Vapnik–Cervonenkis class of sets; for example, see the proof of Lemma A1(a). The covariance kernel of $\nu(\cdot)$ and the pseudo-metric ρ_* are specified below.

Let $c \vee d = \max\{c, d\}$ and $c \wedge d = \min\{c, d\}$.

To prove part (b), we write

$$\begin{aligned}(15.24)\quad Y_i g_{a,b}(X_i) &= \big(U_i + 1\big(X_i \in (\varepsilon_n, 1]\big)\big) \cdot 1\big(X_i \in (a, b]\big) \\ &= U_i 1\big(X_i \in (a, b]\big) + 1\big(X_i \in (a \vee \varepsilon_n, b]\big)\end{aligned}$$

and

$$\begin{aligned}(15.25)\quad E_{F_n} Y_i g_{a,b}(X_i) &= E_{F_n} U_i 1\big(X_i \in (a, b]\big) + P_{F_n}\big(X_i \in (a \vee \varepsilon_n, b]\big) \\ &= P_{F_n}\big(X_i \in (a \vee \varepsilon_n, b]\big) \\ &\to (b - a)/2,\end{aligned}$$

where the second equality uses Assumption CX(b) and the convergence uses Assumption CX(c) and holds by slightly different arguments when $a = 0$ and $a > 0$. Equation (15.25) and $b - a > 0$ imply that $h_{1,n}(g_{a,b}) = n^{1/2} E_{F_n} Y_i g_{a,b}(X_i) \to \infty = h_1(g_{a,b})$ as $n \to \infty$ for all $g_{a,b} \in \mathcal{G}$, which proves part (b).

Part (c) holds because $h_1(g_{a,b}) = \infty$ for all $g_{a,b} \in \mathcal{G}$ and

$$\begin{aligned}(15.26)\quad \inf_{g_{a,b} \in \mathcal{G}} h_{1,n}(g_{a,b}) &= \inf_{g_{a,b} \in \mathcal{G}} n^{1/2} P_{F_n}\big(X_i \in (a \vee \varepsilon_n, b]\big) \\ &= \inf_{a,b : \varepsilon_n \le a < b \le 1} n^{1/2} P_{F_n}\big(X_i \in (a, b]\big) = 0\end{aligned}$$

for all n, where the first equality holds by (15.25) and the last equality holds by Assumption CX(c).

Part (d) holds because $\nu_n(g_{a,b}) + h_{1,n}(g_{a,b}) = O_p(1) + n^{1/2}(b - a)/2 \to_p \infty$ by part (a) and (15.25) for all $g_{a,b} \in \mathcal{G}$. This, combined with Assumption CX(f) (in particular, Assumption S1(d)), proves part (d).

Part (e) holds by part (b) and Assumption CX(f) (in particular, Assumption S2) because $S(\nu(g_{a,b}) + h_1(g_{a,b})) = S(\infty) = 0$ for all $g_{a,b} \in \mathcal{G}$.

To show part (f), we define

$$g_n^*(x) = 1\big(x \in (0, \varepsilon_n]\big). \tag{15.27}$$

Then,

$$h_{1,n}\big(g_n^*\big) = n^{1/2} E_{F_n} Y_i g_n^*(X_i) = P_{F_n}\big(X_i \in (0 \vee \varepsilon_n, \varepsilon_n]\big) = 0 \tag{15.28}$$

for all n, where the second equality holds by (15.25) with $a = 0$ and $b = \varepsilon_n$.
Next, we have

$$\sup_{g_{a,b} \in \mathcal{G}} S\big(\nu_n(g_{a,b}) + h_{1,n}(g_{a,b})\big) \geq S\big(\nu_n\big(g_n^*\big) + h_{1,n}\big(g_n^*\big)\big) = S\big(\nu_n\big(g_n^*\big)\big), \tag{15.29}$$

where the equality holds by (15.28). The asymptotic distribution of $S(\nu_n(g_n^*))$ is established as follows:

$$\begin{aligned} \nu_n\big(g_n^*\big) &= n^{-1/2} \sum_{i=1}^n \big[Y_i 1\big(X_i \in (0, \varepsilon_n]\big) - E_{F_n} Y_i 1\big(X_i \in (0, \varepsilon_n]\big)\big] \\ &= n^{-1/2} \sum_{i=1}^n \big[U_i 1(X_i = \varepsilon_n) + U_i 1\big(X_i \in (0, \varepsilon_n)\big) \\ &\quad + 1\big(X_i \in (\varepsilon_n, 1]\big) 1\big(X_i \in (0, \varepsilon_n]\big) \\ &\quad - E_{F_n} 1\big(X_i \in (\varepsilon_n, 1]\big) 1\big(X_i \in (0, \varepsilon_n]\big)\big] \\ &= n^{-1/2} \sum_{i=1}^n U_i 1(X_i = \varepsilon_n) + n^{-1/2} \sum_{i=1}^n U_i 1\big(X_i \in (0, \varepsilon_n)\big) \\ &\to_d Z^* \sim N(0, 1/2), \end{aligned} \tag{15.30}$$

where the second equality uses $E_{F_n} U_i = 0$ and U_i and X_i are independent. The convergence in distribution in (15.30) holds by a triangular array CLT for the first summand on the second to last line because $U_i 1(X_i = \varepsilon_n)$ has mean zero and variance $E_{F_n} U_i^2 1(X_i = \varepsilon_n) = 1 \cdot P_{F_n}(X_i = \varepsilon_n) = 1/2$ for all n, using Assumption CX(b). The second summand on the second to last line of (15.30) is $o_p(1)$ because its mean is zero and its variance is

$$\begin{aligned} &\operatorname{Var}\left(n^{-1/2} \sum_{i=1}^n U_i 1\big(X_i \in (0, \varepsilon_n)\big)\right) \\ &\quad = \operatorname{Var}\big(U_i 1\big(X_i \in (0, \varepsilon_n)\big)\big) \\ &\quad = E_{F_n} U_i^2 1\big(X_i \in (0, \varepsilon_n)\big) = 1 \cdot P_{F_n}\big(X_i \in (0, \varepsilon_n)\big) = \varepsilon_n/2, \end{aligned} \tag{15.31}$$

where the first equality holds by Assumption CX(d), the second and third equalities hold by Assumption CX(b), and the last equality holds by Assumption CX(c).

Equations (15.29) and (15.30), Assumption S1(d), and the continuous mapping theorem combine to prove part (f).

Part (g) holds if

$$\sup_{g_{a,b}\in\mathcal{G}} S\big(\nu_n(g_{a,b})+h_{1,n}(g_{a,b})\big) \nrightarrow_p 0, \tag{15.32}$$

using part (e). By part (f), for all $\delta \geq 0$,

$$\begin{aligned}\liminf_{n\to\infty} P\Big(\sup_{g_{a,b}\in\mathcal{G}} S\big(\nu_n(g_{a,b})+h_{1,n}(g_{a,b})\big) > \delta\Big) \\ \geq \liminf_{n\to\infty} P\big(S\big(\nu_n\big(g_n^*\big)\big) > \delta\big) \\ = P\big(S\big(Z^*\big) > \delta\big).\end{aligned} \tag{15.33}$$

Now, by the dominated convergence theorem, as $\delta \to 0$,

$$P\big(S\big(Z^*\big) > \delta\big) \to P\big(S\big(Z^*\big) > 0\big) = 1/2, \tag{15.34}$$

where the equality holds because $S(m) > 0$ iff $m < 0$ by Assumption S2 and $P(Z^* < 0) = 1/2$. Hence, the right-hand side in (15.33) is arbitrarily close to $1/2$ for $\delta > 0$ sufficiently small, which implies that (15.32) holds and part (g) is established.

Lastly, we compute the covariance kernel $K(g_{a_1,b_1}, g_{a_2,b_2})$ of the Gaussian process $\nu(\cdot)$. We have

$$\begin{aligned} & E_{F_n} Y_i^2 g_{a_1,b_1}(X_i) g_{a_2,b_2}(X_i) \\ &= E_{F_n}\big(U_i + 1\big(X_i \in (\varepsilon_n, 1]\big)\big)^2 \cdot 1\big(X_i \in (a_1 \vee a_2, b_1 \wedge b_2]\big) \\ &= E_{F_n} U_i^2 1\big(X_i \in (a_1 \vee a_2, b_1 \wedge b_2]\big) \\ &\quad + E_{F_n}(2U_i + 1) 1\big(X_i \in (a_1 \vee a_2 \vee \varepsilon_n, b_1 \wedge b_2]\big) \\ &= P_{F_n}\big(X_i \in (a_1 \vee a_2, b_1 \wedge b_2]\big) + P_{F_n}\big(X_i \in (a_1 \vee a_2 \vee \varepsilon_n, b_1 \wedge b_2]\big) \\ &\to (1/2) 1(a_1 = a_2 = 0) + \max\big\{(b_1 \wedge b_2) - (a_1 \vee a_2), 0\big\} \\ &= K_1(g_{a_1,b_1}, g_{a_2,b_2}),\end{aligned} \tag{15.35}$$

where the third equality uses Assumption CX(b) and the convergence uses Assumption CX(c).

In addition, we have

$$\lim_{n\to\infty} E_{F_n} Y_i g_{a,b}(X_i) = (b-a)/2 = K_2(g_{a,b}), \tag{15.36}$$

where the first equality holds by (15.25). Putting the results of (15.35) and (15.36) together yields

$$
\begin{aligned}
(15.37)\quad & K(g_{a_1,b_1}, g_{a_2,b_2}) \\
&= \lim_{n\to\infty} \big(E_{F_n} Y_i^2 g_{a_1,b_1}(X_i) g_{a_2,b_2}(X_i) \\
&\quad - E_{F_n} Y_i g_{a_1,b_1}(X_i) \cdot E_{F_n} Y_i g_{a_2,b_2}(X_i)\big) \\
&= K_1(g_{a_1,b_1}, g_{a_2,b_2}) - K_2(g_{a_1,b_1}) K_2(g_{a_2,b_2}).
\end{aligned}
$$

The square of the pseudo-metric ρ_* on $\mathcal{G}$ is

$$
\begin{aligned}
(15.38)\quad & \rho_*^2(g_{a_1,b_1}, g_{a_2,b_2}) \\
&= \lim_{n\to\infty} E_{F_n}\big(Y_i g_{a_1,b_1}(X_i) - Y_i g_{a_2,b_2}(X_i) \\
&\quad - E_{F_n} Y_i g_{a_1,b_1}(X_i) + E_{F_n} Y_i g_{a_2,b_2}(X_i)\big)^2.
\end{aligned}
$$

The limit in (15.38) exists and can be computed via calculations analogous to those in (15.25) and (15.35). *Q.E.D.*

PROOF OF THEOREM B5: For notational convenience, we let g denote $g_{a,b}$. By Theorem B4(a), $\nu_n(\cdot) \Rightarrow \nu(\cdot)$ as $n \to \infty$. As in the proof of Theorem 1(a), by an almost sure representation argument (e.g., see Thm. 9.4 of Pollard (1990)), there exist processes $\tilde{\nu}_n(\cdot)$ and $\tilde{\nu}(\cdot)$ on $\mathcal{G}$ that have the same distributions as $\nu_n(\cdot)$ and $\nu(\cdot)$, respectively, for which

$$(15.39)\quad \sup_{g\in\mathcal{G}} \big|\tilde{\nu}_n(g) - \tilde{\nu}(g)\big| \to 0 \quad \text{a.s.}$$

Let $\widetilde{\Omega}$ denote the sample paths for which the convergence in (15.39) holds. By (15.39), $P(\widetilde{\Omega}) = 1$.

For each $\omega \in \widetilde{\Omega}$, we apply the bounded convergence theorem to obtain

$$
\begin{aligned}
(15.40)\quad & \lim_{n\to\infty} \int S\big(\tilde{\nu}_n(g)(\omega) + h_{1,n}(g)\big)\, dQ(g) \\
&= \int S\big(\tilde{\nu}(g)(\omega) + h_1(g)\big)\, dQ(g),
\end{aligned}
$$

which yields the result of the theorem. Now we check the conditions for the bounded convergence theorem. For all $g \in \mathcal{G}$, pointwise convergence holds:

$$S\big(\tilde{\nu}_n(g)(\omega) + h_{1,n}(g)\big) \to S\big(\tilde{\nu}(g)(\omega) + h_1(g)\big) \quad \text{as} \quad n \to \infty$$

by (15.39), Theorem B4(b), and Assumption S1(d). A bound on $S(\tilde{\nu}_n(g)(\omega) + h_{1,n}(g))$ over $g \in \mathcal{G}$ and n sufficiently large is given by $S(\inf_{g^*\in\mathcal{G}} \tilde{\nu}(g^*)(\omega) - \varepsilon)$

for some $\varepsilon > 0$. This follows because, for all $\varepsilon > 0$ and $g \in \mathcal{G}$, we have

$$(15.41)\quad 0 \le S\big(\tilde{\nu}_n(g)(\omega) + h_{1,n}(g)\big) \le S\big(\tilde{\nu}_n(g)(\omega)\big) \le S\Big(\inf_{g^* \in \mathcal{G}} \tilde{\nu}_n\big(g^*\big)(\omega)\Big) \le S\Big(\inf_{g^* \in \mathcal{G}} \tilde{\nu}\big(g^*\big)(\omega) - \varepsilon\Big) < \infty,$$

where the first inequality holds by Assumption S1(c), the second inequality holds by Assumption S1(b) and $h_{1,n}(g) \ge 0$ for all $g \in \mathcal{G}$ by (15.25), the third inequality holds by Assumption S1(b), the fourth inequality holds for all n sufficiently large by (15.39) and Assumption S1(b), and the last inequality holds because $\inf_{g^* \in \mathcal{G}} \tilde{\nu}(g^*)(\omega) > -\infty$ because the sample paths of $\tilde{\nu}(\cdot)$ are bounded a.s. (which follows from $|m(W_i, \theta_0)g(X_i)| \le |m(W_i, \theta_0)| \le |U_i| + 1 < \infty$ a.s. and (15.39)). This completes the proof of (15.40) and the theorem is proved. *Q.E.D.*

15.4. *Proofs of Subsampling Results*

PROOF OF LEMMA B4: For S_1, Assumption SQ(a) holds because (i) if $v \ge 1$, the summand $\sum_{j=p+1}^{k} (\nu_{h_2,j}^2(g)/(h_{2,j,j}(g) + \varepsilon))$ is absolutely continuous for all $g \in \mathcal{G}$, where $\nu_{h_2}(g) = (\nu_{h_2,1}(g), \ldots, \nu_{h_2,k}(g))'$ and $h_{2,j,j}(g)$ denotes the jth diagonal element of $h_2(g)$, (ii) if $v = 0$ and $h_1(g) \ne \infty^p$, the summands $[\nu_{h_2,j}(g) + h_{1,j}(g)]_-^2/(h_{2,j,j}(g) + \varepsilon)$ are absolutely continuous for $x > 0$ and all $j \le p$ such that $h_{1,j}(g) < \infty$, (iii) if $v = 0$ and $h_1(g) = \infty^p$, $S_1(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k) = 0$ and its distribution function equals 1 for all $x > 0$, and (iv) if $S_1(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k)$ is absolutely continuous for all $g \in \mathcal{G}$, then $\int S_1(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k)\, dQ(g)$ is absolutely continuous.

Assumption SQ(b) holds for S_1 because (i) if $v \ge 1$, the summand $\int \sum_{j=p+1}^{k} (\nu_{h_2,j}^2(g)/(h_{2,j,j}(g) + \varepsilon))\, dQ(g)$ has positive density on $[0, \infty)$, and (ii) if $v = 0$ and $h_1(g) \ne \infty^p$ on some $G \subset \mathcal{G}$ such that $Q(G) > 0$, each summand $\int [\nu_{h_2,j}(g) + h_{1,j}(g)]_-^2/(h_{2,j,j}(g) + \varepsilon)\, dQ(g)$ for which $h_{1,j}(g) < \infty$ on some $G \subset \mathcal{G}$ such that $Q(G) > 0$ has positive density on $[0, \infty)$, and so does the sum over $\sum_{j=1}^{p}$.

For S_2, if $v = 0$ and $h_1(g) = \infty^p$ a.s. $[Q]$, then $S_2(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k) = 0$ a.s. $[Q]$, $J_{(h_1,h_2)}(x) = 1$ for all $x > 0$, Assumption SQ(a) holds, and Assumption SQ(b) does not impose any restriction. Otherwise, $v \ge 1$ or $h_1(g) < \infty^p$ on a subset $G \subset \mathcal{G}$ such that $Q(G) > 0$. In this case, the random variable $\int S_2(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k)\, dQ(g)$ has support $[0, \infty)$ and is absolutely continuous. Hence, Assumptions SQ(a)–(b) hold. *Q.E.D.*

The proof of Theorem B6 uses the following lemma.

LEMMA D1: *Suppose Assumptions* M *and* S1 *hold*. *Then*, *for all* $h \in \mathcal{H}$, *under any sequence* $\{(\theta_n, F_n) : n \geq 1\} \in \mathrm{Seq}^b(h_1^*, h)$,

$$T_n(\theta_n) \to_d \int S\big(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k\big)\, dQ(g) \\ \sim J_{(h_1,h_2)} \quad as \quad n \to \infty.$$

COMMENT: Condition (iv) of $\mathrm{Seq}^b(h_1^*, h)$ is not needed for the result of Lemma D1 to hold.

PROOF OF THEOREM B6: First, we prove part (a). Suppose $\{(\theta_n, F_n) : n \geq 1\} \in \mathrm{Seq}^b$. Then, there exist $h \in \mathcal{H}$ and $h_1^* \in \mathcal{H}_1^*(h)$ such that $\{(\theta_n, F_n) : n \geq 1\} \in \mathrm{Seq}^b(h_1^*, h)$. We need to show that, under $\{(\theta_n, F_n) : n \geq 1\}$, $\limsup_{n\to\infty} P_{F_n}(T_n(\theta_n) \leq c_{n,b}(\theta_n, 1-\alpha)) \geq 1-\alpha$. The asymptotic distribution of $T_n(\theta_n)$ is given by Lemma D1. We now determine the probability limit of $c_{n,b}(\theta_n, 1-\alpha)$.

Let $J_{(h_1,h_2)}(x)$ for $x \in R$ denote the distribution function of $J_{(h_1,h_2)}$. By Lemma 5 in Andrews and Guggenberger (2010), if (i) $U_{n,b}(\theta_n, x) \to_p J_{(h_1^*,h_2)}(x)$ for all $x \in C(J_{(h_1^*,h_2)})$, where $C(J_{(h_1^*,h_2)})$ denotes the continuity points of $J_{(h_1^*,h_2)}$, and (ii) for all $\xi > 0$, $J_{(h_1^*,h_2)}(c_\infty + \xi) > 1-\alpha$, where c_∞ is the $1-\alpha$ quantile of $J_{(h_1^*,h_2)}$, then

$$c_{n,b}(\theta_n, 1-\alpha) \to_p c_\infty. \tag{15.42}$$

Condition (i) holds by the properties of U-statistics of degree b and $T_{n,b,j}(\theta_n) \to_d J_{(h_1^*,h_2)}$ (see Thm. 2.1(i) in Politis and Romano (1994)). The latter holds by Lemma D1 because subsample j is an i.i.d. sample of size b from the population distribution.

By Assumption S1(c), $J_{(h_1,h_2)}(x) = 0$ $\forall x < 0$ for $h \in \mathcal{H}$. Thus, $c_\infty \geq 0$. If $v = 0$ and $h_1(g) = \infty^p$ a.s. $[Q]$, then $J_{(h_1^*,h_2)}(0) = 1$, $c_\infty = 0$, $J_{(h_1^*,h_2)}(c_\infty + \xi) = 1 > 1-\alpha$. In all other cases, Assumption SQ(b) applies, $J_{(h_1^*,h_2)}(0) < 1$, and $J_{(h_1^*,h_2)}(c_\infty + \xi) > J_{(h_1^*,h_2)}(c_\infty) \geq 1-\alpha$. Thus, condition (ii) holds and (15.42) is established.

If $c_\infty > 0$, $c_\infty \in C(J_{(h_1,h_2)})$ by Assumption SQ(a). Thus,

$$\liminf_{n\to\infty} P_{F_n}\big(T_n(\theta_n) \leq c_{n,b}(\theta_n, 1-\alpha)\big) \\ = J_{(h_1,h_2)}(c_\infty) \geq J_{(h_1^*,h_2)}(c_\infty) = 1-\alpha, \tag{15.43}$$

where the first equality holds by (15.42) and Lemma D1, the inequality holds by Assumption S1(b) and $h_1^* \leq h_1$, and the second equality holds by Assumption SQ(a) and the definition of c_∞.

If $c_\infty = 0$, for some set $G \subset \mathcal{G}$ with $Q(G) = 1$, we have

$$\begin{aligned}
(15.44)\quad & P_{F_n}\big(T_n(\theta_n) \le c_{n,b}(\theta_n, 1-\alpha)\big) \\
&\ge P_{F_n}\big(T_n(\theta_n) \le 0\big) \\
&= P_{F_n}\Bigg(\frac{n^{1/2}\bar{m}_{n,j}(\theta_n, g)}{\bar{\sigma}_{n,j}(\theta_n, g)} \ge 0\ \forall j \le p\ \&\ \frac{\bar{m}_{n,j}(\theta_n, g)}{\bar{\sigma}_{n,j}(\theta_n, g)} = 0 \\
&\qquad \forall j = p+1, \ldots, k, \forall g \in G\Bigg) \\
&\to P\Bigg(\frac{\nu_{h,j}(g) + h_{1,j}(g)}{h_{2,j,j}(g) + \varepsilon} \ge 0\ \forall j \le p\ \&\ \frac{\nu_{h,j}(g)}{h_{2,j,j}(g) + \varepsilon} = 0 \\
&\qquad \forall j = p+1, \ldots, k, \forall g \in G\Bigg) \\
&= P\big(S\big(\nu_h(g) + h_1(g), h_2(g) + \varepsilon I_k\big) = 0\ \forall g \in G\big) \\
&= J_{(h_1,h_2)}(0) \ge J_{(h_1^*,h_2)}(0) \ge 1-\alpha,
\end{aligned}$$

where $\bar{\sigma}_{n,j}(\theta, g)$ and $h_{2,j,j}(g)$ denote the jth diagonal elements of $\bar{\Sigma}_n(\theta, g)$ and $h_2(g)$, respectively. In (15.44), the first inequality holds because $c_{n,b}(\theta_n, 1-\alpha)$ is the $1-\alpha$ sample quantile of the subsample test statistics and the test statistics are nonnegative (by Assumption S1(a)), the first and second equalities hold by Assumption S2, the convergence holds by Lemma A1(a)–(b), the third equality holds by the definition of $J_{(h_1,h_2)}$, and the last inequality holds because 0 is the $1-\alpha$ quantile of $J_{(h_1^*,h_2)}$.

Next, we prove part (b). Let $(\theta_n^*, F_n^*) = (\theta, F)$ for $n \ge 1$, where (θ, F) is specified in Assumption C. Then, $\{(\theta_n^*, F_n^*) : n \ge 1\} \in \mathrm{Seq}^b(h_1^*, h)$, where $h_1^* = h_{1,F}(\theta)$ and $h = (h_{1,F}(\theta), h_{2,F}(\theta))$. Thus,

$$\begin{aligned}
(15.45)\quad & \liminf_{n\to\infty} P_{F_n^*}\big(T_n\big(\theta_n^*\big) \le c_{n,b}\big(\theta_n^*, 1-\alpha\big)\big) \\
&= J_{(h_1,h_2)}(c_\infty) = J_{(h_1^*,h_2)}(c_\infty) = 1-\alpha.
\end{aligned}$$

This and the result of Theorem B6(a) establish part (b).

Lastly, we prove part (c). Suppose Assumption Sub holds and $\{(\theta_{m_n}, F_{m_n}) : n \ge 1\}$ belongs to Seq^b (where Seq^b is defined with m_n in place of n). Then,

$$\begin{aligned}
(15.46)\quad \mathrm{AsyCS} &= \lim_{n\to\infty} P_{F_{m_n}}\big(T_n(\theta_{m_n}) \le c_{n,b}(\theta_{m_n}, 1-\alpha)\big) \\
&\ge \inf_{\{(\theta_n,F_n):n\ge 1\}\in \mathrm{Seq}^b} \liminf_{n\to\infty} P_{F_n}\big(T_n(\theta_n) \le c_{n,b}(\theta_n, 1-\alpha)\big) \\
&= 1-\alpha
\end{aligned}$$

using Theorem B6(b). On the other hand,

$$\begin{aligned}(15.47)\quad \mathrm{AsyCS} &= \liminf_{n\to\infty} \inf_{(\theta,F)\in\mathcal{F}} P_F\big(T_n(\theta) \le c_{n,b}(\theta, 1-\alpha)\big) \\ &\le \inf_{\{(\theta_n,F_n):n\ge 1\}\in \mathrm{Seq}^b} \liminf_{n\to\infty} P_{F_n}\big(T_n(\theta_n) \le c_{n,b}(\theta_n, 1-\alpha)\big) \\ &= 1-\alpha.\end{aligned}$$

Thus, we have $\mathrm{AsyCS} = 1-\alpha$. *Q.E.D.*

PROOF OF LEMMA D1: By the same argument as used above to show (14.20), but with $\nu_{\widehat{h}_{2,n}(\theta_c)}(g)$ and $\varphi_n(\theta_c, g)$ replaced by $\nu_{n,F_n}(\theta_n, g)$ and $h_{1,n,F_n}(\theta_n, g)$, respectively, we have

$$(15.48)\quad T_n(\theta_n) \to_d T(h) = \int S\big(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k\big)\, dQ(g),$$

where $\nu_{n,F_n}(\theta_n, \cdot) \Rightarrow \nu_{h_2}(\cdot)$ by Lemma A1(a), $h_{1,n,F_n}(\theta_n, g) \to h_1(g)\ \forall g \in \mathcal{G}$ by Definition $\mathrm{Seq}^b(h_1^*, h)$(ii), and $d(\widehat{h}_{2,n}(\theta_n), h_2) \to 0$ by Lemma A1(b) and (12.26). Note that the assumption that $\{(\theta_n, F_n): n \ge 1\}$ satisfies Definition $\mathrm{Seq}^b(h_1^*, h)$ and Assumption M implies that $\{(\theta_n, F_n): n \ge 1\}$ satisfies Definition $\mathrm{SubSeq}(h_2)$ and hence the conditions of Lemma A1 hold. *Q.E.D.*

16. SUPPLEMENTAL APPENDIX E

This appendix proves Lemma A1, which is stated in Supplemental Appendix A.

16.1. *Preliminary Lemmas E1–E3*

Before we prove Lemma A1, we review a few concepts from Pollard (1990) and state several lemmas that are used in the proof.

DEFINITION E1—Pollard (1990, Definition 3.3): The packing number $D(\xi, \rho, G)$ for a subset G of a metric space $(\mathcal{G}, \rho)$ is defined as the largest b for which there exist points $g^{(1)}, \ldots, g^{(b)}$ in G such that $\rho(g^{(s)}, g^{(s')}) > \xi$ for all $s \ne s'$. The covering number $N(\xi, \rho, G)$ is defined to be the smallest number of closed balls with ρ-radius ξ whose union covers G.

It is easy to see that $N(\xi, \rho, G) \le D(\xi, \rho, G) \le N(\xi/2, \rho, G)$.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be the underlying probability space equipped with probability distribution $\mathbf{P}$. Let $\{f_{n,i}(\omega, g): g \in \mathcal{G}, i \le n, n \ge 1\}$ be a triangular array of random processes. Let

$$(16.1)\quad \mathcal{F}_{n,\omega} = \big\{\big(f_{n,1}(\omega, g), \ldots, f_{n,n}(\omega, g)\big)' : g \in \mathcal{G}\big\}.$$

Because $\mathcal{F}_{n,\omega} \subset R^n$, we use the Euclidean metric $\|\cdot\|$ on this space. For simplicity, we omit the metric argument in the packing number function, that is, we write $D(\xi, G)$ in place of $D(\xi, \|\cdot\|, G)$ when $G \subset \mathcal{F}_{n,\omega}$.

Let $\odot$ denote the element-by-element product. For example, for $a, b \in R^n$, $a \odot b = (a_1 b_1, \ldots, a_n b_n)'$. Let *envelope functions* of a triangular array of processes $\{f_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ be an array of functions $\{F_n(\omega) = (F_{n,1}(\omega), \ldots, F_{n,n}(\omega))' : n \geq 1\}$ such that $|f_{n,i}(\omega, g)| \leq F_{n,i}(\omega)$ $\forall i \leq n, n \geq 1, g \in \mathcal{G}, \omega \in \Omega$.

DEFINITION E2—Pollard (1990, Definition 7.9): A triangular array of processes $\{f_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ is said to be manageable with respect to envelopes $\{F_n(\omega) : n \geq 1\}$ if there exists a deterministic real function λ on $(0, 1]$ for which (i) $\int_0^1 \sqrt{\log \lambda(\xi)}\, d\xi < \infty$ and (ii) $D(\xi \|\alpha \odot F_n(\omega)\|, \alpha \odot \mathcal{F}_{n,\omega}) \leq \lambda(\xi)$ for $0 < \xi \leq 1$, all $\omega \in \Omega$, all n-vectors α of nonnegative weights, and all $n \geq 1$.

LEMMA E1: *If a row-wise i.i.d. triangular array of random processes $\{\phi_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ is manageable with respect to the envelopes $\{F_n(\omega) : n \geq 1\}$ and $c_n(\omega) = (c_{n,1}(\omega), \ldots, c_{n,n}(\omega))'$ is an R^n-valued function on the underlying probability space, then*

(a) $\{\phi_{n,i}(\omega, g) c_{n,i}(\omega) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ *is manageable with respect to the envelopes*

$$F_n(\omega) = \big(F_{n,1}(\omega)\big|c_{n,1}(\omega)\big|, \ldots, F_{n,n}(\omega)\big|c_{n,n}(\omega)\big|\big)' \quad \textit{for} \quad n \geq 1, \tag{16.2}$$

(b) $\{E\phi_{n,i}(\cdot, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ *is manageable with respect to the envelopes $\{EF_n : n \geq 1\}$ provided $EF_{n,1} < \infty$ for all $n \geq 1$, and*

(c) *if another triangular array of random processes $\{\phi^*_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ is manageable with respect to the envelopes $\{F^*_n(\omega) : n \geq 1\}$, then $\{\phi^*_{n,i}(\omega, g) + \phi_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ is manageable with respect to the envelopes $\{F_n(\omega) + F^*_n(\omega) : n \geq 1\}$.*

LEMMA E2: *If the triangular array of processes $\{f_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ is manageable with respect to the envelopes $\{F_n(\omega) = (F_{n,1}(\omega), \ldots, F_{n,n}(\omega))' : n \geq 1\}$, and there exist $0 < \eta < 1$ and $0 < B^* < \infty$ such that $n^{-1} \sum_{i \leq n} EF^{1+\eta}_{n,i} \leq B^*$ for all $n \geq 1$, then*

$$\sup_{g \in \mathcal{G}} \left| n^{-1} \sum_{i=1}^n \big(f_{n,i}(\omega, g) - Ef_{n,i}(\cdot, g)\big) \right| \to_p 0. \tag{16.3}$$

Lemma E1(b)–(c) imply that if $\{f_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ is manageable, then the triangular array of recentered processes $\{f_{n,i}(\omega, g) - Ef_{n,i}(\cdot, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ also is manageable with respect to their corresponding envelopes. Lemma E2 is a uniform weak law of large numbers for triangular arrays of row-wise independent random processes. Lemma E2 is a complement to Theorem 8.2 in Pollard (1990), which is a uniform weak law of large numbers for independent sequences of random processes.

Lemma A1(a) is a functional central limit theorem result for multidimensional empirical processes. We prove it using a functional central limit theorem for real-valued empirical processes given in Pollard (1990, Thm. 10.3) and the Cramér–Wold device.

For $a \in R^k/\{0_k\}$, let

$$f_{n,i}(\omega, g) = a' D_{F_n}^{-1/2}(\theta_n) n^{-1/2} \times \big[m\big(W_{n,i}(\omega), \theta_n, g\big) - E_{F_n} m\big(W_{n,i}(\cdot), \theta_n, g\big)\big] \quad \text{for} \quad \omega \in \Omega, \quad g \in \mathcal{G}, \tag{16.4}$$

where $W_{n,i}(\cdot) = W_i$, and the index n in $W_{n,i}$ signifies the fact that the distribution of W_i is changing with n. The random variable $f_{n,i}(\omega, g)$ depends on a, but for notational simplicity, a does not appear explicitly in $f_{n,i}(\omega, g)$. By definition, we have

$$a'\nu_{n,F_n}(\theta_n, g) = \sum_{i=1}^{n} f_{n,i}(\omega, g). \tag{16.5}$$

Let

$$\rho_{n,a}\big(g, g^*\big) = \big(nE\big|f_{n,i}(\cdot, g) - f_{n,i}\big(\cdot, g^*\big)\big|^2\big)^{1/2} \quad \text{for} \quad g, g^* \in \mathcal{G}. \tag{16.6}$$

We show in the proof of Lemma E3 below that, under the assumptions, the sequence $\{\rho_{n,a}(g, g^*) : n \geq 1\}$ converges for each pair $g, g^* \in \mathcal{G}$. In consequence, the pointwise limit of $\rho_{n,a}(\cdot, \cdot)$ is an appropriate choice for the pseudo-metric on $\mathcal{G}$. Denote the limit by $\rho_a(\cdot, \cdot)$, that is,

$$\rho_a\big(g, g^*\big) = \lim_{n \to \infty} \rho_{n,a}\big(g, g^*\big). \tag{16.7}$$

LEMMA E3: *For all $a \in R^k \setminus \{0\}$ and any subsequence $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in$* SubSeq$(h_2)$, *for some $k \times k$-matrix-valued covariance kernel h_2 on $\mathcal{G} \times \mathcal{G}$,*

(a) *$\mathcal{G}$ is totally bounded under the pseudo-metric ρ_a,*

(b) *the finite-dimensional distributions of $a'\nu_{a_n,F_{a_n}}(\theta_{a_n}, g)$ have Gaussian limits with zero means and covariances given by $a'h_2(g, g^*)a$ $\forall g, g^* \in \mathcal{G}$, which uniquely determine a Gaussian distribution ν_a concentrated on the space of uniformly $\rho_a(\cdot, \cdot)$-continuous bounded functionals on $\mathcal{G}$, $U_{\rho_a}(\mathcal{G})$, and*

(c) *$a'\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot)$ converges in distribution to ν_a.*

The proofs of Lemmas E1–E3 are given below. The proof of Lemma E2 uses the maximal inequality in (7.10) of Pollard (1990). The proof of Lemma E3 uses the real-valued empirical process result of Theorem 10.6 in Pollard (1990).

16.2. *Proof of Lemma A1(a)*

Lemma A1 is stated in terms of subsequences $\{a_n\}$. For notational simplicity, we prove it for the sequence $\{n\}$. All of the arguments in this subsection and the next go through with $\{a_n\}$ in place of $\{n\}$.

The following three conditions are sufficient for weak convergence: (a) $(\mathcal{G}, \rho)$ is a totally bounded pseudo-metric space, (b) finite-dimensional convergence holds: $\forall\{g^{(1)}, \dots, g^{(L)}\} \subset \mathcal{G}$, $(\nu_{n,F_n}(\theta_n, g^{(1)})', \dots, \nu_{n,F_n}(\theta_n, g^{(L)})')'$ converges in distribution, and (c) $\{\nu_{n,F_n}(\theta_n, \cdot) : n \geq 1\}$ is stochastically equicontinuous. (For example, see Thm. 10.2 of Pollard (1990).)

First, we establish the total boundedness of the pseudo-metric space $(\mathcal{G}, \rho)$, that is, $N(\xi, \rho, \mathcal{G}) < \infty$ for all $\xi > 0$. This is done by constructing a finite collection of closed balls that covers $(\mathcal{G}, \rho)$.

Consider $\xi > 0$. Let $B_\rho(g, \xi)$ denote a closed ball centered at g with ρ-radius ξ. Let $\#G$ denote the number of elements in G when G is a finite set. (Throughout this proof, G denotes a subset of $\mathcal{G}$, not the envelope function that appears in Assumption M.) For $j = 1, \dots, k$, let e_j be a k-dimensional vector with the jth coordinate equal to 1 and all other coordinates equal to zero. Then, $e_j \in R^k \setminus \{0\}$, and by Lemma E3(a), the pseudo-metric spaces $(\mathcal{G}, \rho_{e_j})$ are totally bounded. Consequently, for all $G \subset \mathcal{G}$, (G, ρ_{e_j}) is totally bounded. Our construction of the collection of closed balls is based on the following relationship between $\{\rho_{e_j} : j \leq k\}$ and ρ: $\forall g, g^* \in \mathcal{G}$,

$$
\begin{aligned}
(16.8)\qquad \rho^2(g, g^*) &= \operatorname{tr}\big(h_2(g, g) - h_2(g, g^*) - h_2(g^*, g) + h_2(g^*, g^*)\big) \\
&= \lim_{n\to\infty} E_{F_n} \big\| D_{F_n}^{-1/2}(\theta_n)\big[\widetilde{m}(W_i, \theta_n, g) - \widetilde{m}(W_i, \theta_n, g^*)\big]\big\|^2 \\
&= \lim_{n\to\infty} \sum_{j=1}^{k} \rho_{n,e_j}^2(g, g^*) = \sum_{j=1}^{k} \rho_{e_j}^2(g, g^*),
\end{aligned}
$$

where the second equality holds by (16.7), which is proved in (16.40)–(16.41).

We start with $j = 1$. Because $(\mathcal{G}, \rho_{e_1})$ is totally bounded, we can find a set $G_1 \subset \mathcal{G}$ such that

$$
(16.9)\qquad \#G_1 = N(\xi_k, \rho_{e_1}, \mathcal{G}) \quad \text{and} \quad \sup_{g\in\mathcal{G}} \min_{g^*\in G_1} \rho_{e_1}(g, g^*) \leq \xi_k,
$$

where $\xi_k = \xi/(2\sqrt{k})$. For all $g \in G_1$, let $B^1_{\rho_{e_1}}(g, \xi_k) = B_{\rho_{e_1}}(g, \xi_k) \cap \mathcal{G}$. Then, $\bigcup_{g\in G_1} B^1_{\rho_{e_1}}(g, \xi_k)$ covers $\mathcal{G}$.

Because $B^1_{\rho_{e_1}}(g, \xi_k) \subset \mathcal{G}$, $(B^1_{\rho_{e_1}}(g, \xi_k), \rho_{e_2})$ is totally bounded. We are then able to choose a set $G_{2,g}$ such that

$$
\begin{aligned}
(16.10)\qquad &\#G_{2,g} = N\big(\xi_k, \rho_{e_2}, B^1_{\rho_{e_1}}(g, \xi_k)\big) \quad \text{and} \\
&\sup_{g'\in B^1_{\rho_{e_1}}(g,\xi_k)} \min_{g^*\in G_{2,g}} \rho_{e_2}(g', g^*) \leq \xi_k.
\end{aligned}
$$

Let $G_2 = \bigcup_{g\in G_1} G_{2,g}$. We have $\#G_2 = \sum_{g\in G_1} \#G_{2,g} < \infty$. For all $g \in G_1$ and $g' \in G_{2,g}$, let

$$B^2_{\rho_{e_2}}(g', \xi_k) = B_{\rho_{e_2}}(g', \xi_k) \cap B^1_{\rho_{e_1}}(g, \xi_k). \tag{16.11}$$

By construction, $\bigcup_{g'\in G_{2,g}} B^2_{\rho_{e_2}}(g', \xi_k)$ covers $B^1_{\rho_{e_1}}(g, \xi_k)$. Because $\bigcup_{g\in G_1} B^1_{\rho_{e_1}}(g, \xi_k)$ covers $\mathcal{G}$, $\bigcup_{g'\in G_2} B^2_{\rho_{e_2}}(g', \xi_k)$ covers $\mathcal{G}$.

Repeat the previous steps to obtain, in turn, G_3, $\{B^3_{\rho_{e_3}}(g, \xi_k) : g \in G_3\}, \ldots, G_k, \{B^k_{\rho_{e_k}}(g, \xi_k) : g \in G_k\}$. One can induce that (i) $\#G_k < \infty$, (ii) $\bigcup_{g'\in G_k} B^k_{\rho_{e_k}}(g', \xi_k)$ covers $\mathcal{G}$, and (iii) $\forall g \in \mathcal{G}$, there exists $(g^{(k)}, g^{(k-1)}, \ldots, g^{(1)}) \in G_k \times G_{k-1} \times \cdots \times G_1$ such that

$$g \in B^k_{\rho_{e_k}}\big(g^{(k)}, \xi_k\big) \subset B^{k-1}_{\rho_{e_{k-1}}}\big(g^{(k-1)}, \xi_k\big) \subset \cdots \subset B^1_{\rho_{e_1}}\big(g^{(1)}, \xi_k\big). \tag{16.12}$$

Thus,

$$\rho\big(g, g^{(k)}\big) = \left(\sum_{j=1}^{k} \rho^2_{e_j}\big(g, g^{(k)}\big)\right)^{1/2} \le \left(\frac{\xi^2}{4k} + \frac{4\xi^2}{4k} + \cdots + \frac{4\xi^2}{4k}\right)^{1/2} < \xi. \tag{16.13}$$

Equation (16.13) implies that $\bigcup_{g\in G_k} B^k_{\rho}(g, \xi)$ covers $\mathcal{G}$, G_k is the desired finite collection we set out to construct, $N(\xi, \rho, \mathcal{G}) \le \#G_k < \infty$, and $(\mathcal{G}, \rho)$ is totally bounded.

Second, we show that finite-dimensional convergence holds. By Lemma E3, the finite-dimensional random vector $(a'\nu_{n,F_n}(\theta_n, g^{(1)}), \ldots, a'\nu_{n,F_n}(\theta_n, g^{(L)}))'$ converges in distribution:

$$\begin{pmatrix} a'\nu_{n,F_n}\big(\theta_n, g^{(1)}\big) \\ \vdots \\ a'\nu_{n,F_n}\big(\theta_n, g^{(L)}\big) \end{pmatrix} \to_d N\left(0, \begin{pmatrix} a'h_2\big(g^{(1)}, g^{(1)}\big)a & \cdots & a'h_2\big(g^{(1)}, g^{(L)}\big)a \\ \vdots & \cdots & \vdots \\ a'h_2\big(g^{(L)}, g^{(1)}\big)a & \cdots & a'h_2\big(g^{(L)}, g^{(L)}\big)a \end{pmatrix}\right) \tag{16.14}$$

for all $a \in R^k$. Thus, by the Cramér–Wold device, for all $g^{(1)}, g^{(2)}, \ldots, g^{(L)} \in \mathcal{G}$,

$$\begin{pmatrix} \nu_{n,F_n}\big(\theta_n, g^{(1)}\big) \\ \vdots \\ \nu_{n,F_n}\big(\theta_n, g^{(L)}\big) \end{pmatrix} \to_d N\left(0, \begin{pmatrix} h_2\big(g^{(1)}, g^{(1)}\big) & \cdots & h_2\big(g^{(1)}, g^{(L)}\big) \\ \vdots & \cdots & \vdots \\ h_2\big(g^{(L)}, g^{(1)}\big) & \cdots & h_2\big(g^{(L)}, g^{(L)}\big) \end{pmatrix}\right). \tag{16.15}$$

Lastly, we show that $\{\nu_{n,F_n}(\theta_n, \cdot) : n \ge 1\}$ is stochastically equicontinuous with respect to ρ. By Lemma E3, $\{e_j'\nu_{n,F_n}(\theta_n, \cdot) : n \ge 1\}$ is stochastically equicontinuous with respect to ρ_{e_j} for all $j \leq k$. (Weak convergence implies stochastic equicontinuity.) Because $\rho(g, g^*) \geq \rho_{e_j}(g, g^*)$ for all $g, g^* \in \mathcal{G}$, $\{e_j'\nu_{n,F_n}(\theta_n, \cdot) : n \geq 1\}$ is stochastically equicontinuous with respect to ρ for all $j \leq k$. Note that $e_j'\nu_{n,F_n}(\theta_n, \cdot)$ is the jth coordinate of $\nu_{n,F_n}(\theta_n, \cdot)$. Therefore, $\{\nu_{n,F_n}(\theta_n, \cdot) : n \geq 1\}$ is stochastically equicontinuous with respect to ρ. *Q.E.D.*

16.3. *Proof of Lemma A1(b)*

It suffices to show that each element of $D_F^{-1/2}(\theta)\widehat{\Sigma}_n(\theta, g, g^*)D_F^{-1/2}(\theta)$ converges in probability uniformly over $g, g^* \in \mathcal{G}$. Suppose $1 \leq j, j' \leq k$. The (j, j')th element of $D_{F_n}^{-1/2}(\theta_n)\widehat{\Sigma}_n(\theta_n, g, g^*)D_{F_n}^{-1/2}(\theta_n)$ can be decomposed into two parts:

$$
\begin{aligned}
(16.16)\quad & n^{-1}\sum_{i=1}^{n}\sigma_{F_n,j}^{-1}(\theta_n)m_j(W_i, \theta_n)m_{j'}(W_i, \theta_n)\sigma_{F_n,j'}^{-1}(\theta_n)g_j(X_i)g_{j'}^*(X_i) \\
&\quad - \sigma_{F_n,j}^{-1}(\theta_n)\bar{m}_{n,j}(\theta_n, g)\bar{m}_{n,j'}\big(\theta_n, g^*\big)\sigma_{F_n,j'}^{-1}(\theta_n) \\
&\equiv n^{-1}\sum_{i=1}^{n} f_{n,i,j,j'}^{mm}\big(\omega, g, g^*\big) \\
&\quad - n^{-1}\sum_{i=1}^{n} f_{n,i,j}^{m}(\omega, g)\left(n^{-1}\sum_{i=1}^{n} f_{n,i,j'}^{m}\big(\omega, g^*\big)\right),
\end{aligned}
$$

where

$$
\begin{aligned}
(16.17)\quad & f_{n,i,j}^{m}(\omega, g) = \sigma_{F_n,j}^{-1}(\theta_n)m_j(W_i, \theta_n)g_j(X_i) \quad \text{and} \\
& f_{n,i,j,j'}^{mm}\big(\omega, g, g^*\big) = f_{n,i,j}^{m}(\omega, g)f_{n,i,j'}^{m}\big(\omega, g^*\big).
\end{aligned}
$$

Note that $\{f_{n,i,j,j'}^{mm}(\omega, g, g^*) : g, g^* \in \mathcal{G}, i \leq n, n \geq 1\}$ and $\{f_{n,i,j}^{m}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ are triangular arrays of row-wise i.i.d. random processes. We show the uniform convergence of their sample means using Lemma E2.

We first study $f_{n,i,j}^{m}(\omega, g)$. Let

$$
(16.18)\quad \mathcal{F}_{n,\omega,j}^{m} = \big\{\big(f_{n,1,j}^{m}(\omega, g), \ldots, f_{n,n,j}^{m}(\omega, g)\big)' : g \in \mathcal{G}\big\}.
$$

By Assumption M(c) and Lemma E1, $\{f_{n,i,j}^{m}(\omega, g) : i \leq n,\ g \in \mathcal{G}\}$ are manageable with respect to the envelopes

$$
\begin{aligned}
(16.19)\quad & F_{n,\cdot,j}^{m}(\omega) = \big(F_{n,1,j}^{m}(\omega), \ldots, F_{n,n,j}^{m}(\omega)\big)', \quad \text{where} \\
& F_{n,i,j}^{m}(\omega) = G(X_i)\sigma_{F_n,j}^{-1}(\theta_n)\big|m_j(W_i, \theta_n)\big|.
\end{aligned}
$$

In consequence, there exist functions $\lambda_j : (0,1] \to [0,\infty)$ for $j \le k$ such that

$$(16.20)\quad D\big(\xi \big| \alpha \odot F^m_{n,\cdot,j}\big|, \alpha \odot \mathcal{F}^m_{n,\omega,j}\big) \le \lambda_j(\xi)$$

for all $\alpha \in [0,\infty)^n$, $\omega \in \Omega$, and $n \ge 1$ and $\sqrt{\log \lambda_j(\xi)}$ is integrable over $(0,1]$.

Because the data are i.i.d., we have, for all $0 < \eta \le 1$ and all n,

$$(16.21)\quad n^{-1}\sum_{i=1}^{n} E\big(F^m_{n,i,j}\big)^{1+\eta} = E\big(F^m_{n,1,j}\big)^{1+\eta}$$
$$\le \big(E_{F_n} G^{\delta_1}(X_i)\big)^{(1+\eta)/\delta_1} \left(E_{F_n}\left|\frac{m_j(W_1,\theta_n)}{\sigma_{F_{n,h},j}(\theta_n)}\right|^{\delta_2}\right)^{(1+\eta)/\delta_2}$$
$$< \infty,$$

where $\delta_2 = (1+\eta)\delta_1/(\delta_1 - 1 - \eta)$. The first inequality above holds by Hölder's inequality and the second holds by Assumption M(b), $\delta_2 \le 2 + 4/(\delta_1 - 1 - \eta) \le 2 + 4/(4\delta^{-1} + 1 - \eta) \le 2 + \delta$, and condition (vi) of (2.3). Therefore, by Lemma E2,

$$(16.22)\quad \sup_{g\in\mathcal{G}}\left| n^{-1}\sum_{i=1}^{n} f^m_{n,i,j}(\omega, g) - Ef^m_{n,1,j}(\cdot, g)\right| \to_p 0.$$

Now we study $f^{mm}_{n,i,j,j'}(\omega, g, g^*)$. For all $n \ge 1$ and $\omega \in \Omega$, let

$$(16.23)\quad \mathcal{F}^{mm}_{n,\omega,j,j'} = \big\{\big(f^{mm}_{n,1,j,j'}(\omega, g, g^*), \ldots, f^{mm}_{n,n,j,j'}(\omega, g, g^*)\big)' : g, g^* \in \mathcal{G}\big\}.$$

Then, $\mathcal{F}^{mm}_{n,\omega,j,j'} = \mathcal{F}^m_{n,\omega,j} \odot \mathcal{F}^m_{n,\omega,j}$. Let $F^{mm}_{n,\cdot,j,j'}(\omega) = F^m_{n,\cdot,j}(\omega) \odot F^m_{n,\cdot,j'}(\omega)$. We have: for all $\alpha \in [0,\infty)^n$, $\omega \in \Omega$, and $n \ge 1$,

$$\begin{aligned}(16.24)\quad & D\big(\xi \big| \alpha \odot F^{mm}_{n,\cdot,j,j'}(\omega)\big|, \alpha \odot \mathcal{F}^{mm}_{n,\omega,j,j'}\big) \\ &= D\big(\xi \big| \alpha \odot F^{mm}_{n,\cdot,j,j'}(\omega)\big|, \alpha \odot \mathcal{F}^m_{n,\omega,j} \odot \mathcal{F}^m_{n,\omega,j}\big) \\ &\le D\left(\frac{\xi}{4}\big|\alpha \odot F^m_{n,\cdot,j'}(\omega) \odot F^m_{n,\cdot,j}(\omega)\big|, \alpha \odot F^m_{n,\cdot,j'}(\omega) \odot \mathcal{F}^m_{n,\omega,j}\right) \\ &\quad \cdot D\left(\frac{\xi}{4}\big|\alpha \odot F^m_{n,\cdot,j}(\omega) \odot F^m_{n,\cdot,j'}(\omega)\big|, \alpha \odot F^m_{n,\cdot,j}(\omega) \odot \mathcal{F}^m_{n,\omega,j'}\right) \\ &\le \lambda_j(\xi/4)\lambda_{j'}(\xi/4),\end{aligned}$$

where the first inequality holds by equation (5.2) in Pollard (1990) and the second inequality holds by (16.20). We have

$$
\begin{aligned}
(16.25)\quad & \int_0^1 \sqrt{\log\bigl(\lambda_j(\xi/4)\lambda_{j'}(\xi/4)\bigr)}\,d\xi \\
&= \int_0^1 \sqrt{\log \lambda_j(\xi/4) + \log \lambda_{j'}(\xi/4)}\,d\xi \\
&\leq 4\int_0^{1/4} \bigl(\sqrt{\log \lambda_j(\xi)} + \sqrt{\log \lambda_{j'}(\xi)}\bigr)\,d\xi < \infty,
\end{aligned}
$$

where the first inequality holds by $\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}$. Therefore, $\{f^{mm}_{n,i,j,j'}(\omega, g, g^*) : g, g^* \in \mathcal{G}, i \leq n, n \geq 1\}$ are manageable with respect to the envelopes $\{F^{mm}_{n,\cdot,j,j'}(\omega) : n \geq 1\}$.

Let η be a small positive number. We have

$$
\begin{aligned}
(16.26)\quad & n^{-1}\sum_{i\leq n} E\bigl(F^{mm}_{n,j,j'}(\cdot)\bigr)^{1+\eta} \\
&= E\bigl(F^{mm}_{n,j,j'}(\cdot)\bigr)^{1+\eta} \\
&\leq \bigl[E_{F_n} G^{\delta_3}(X_1)\bigr]^{2(1+\eta)/\delta_3} \left[E_{F_n}\left|\frac{m_j(W_1,\theta_n)}{\sigma_{F_n,j}(\theta_n)}\right|^{2+\delta}\right]^{(1+\eta)/(2+\delta)} \\
&\quad \times \left[E_{F_n}\left|\frac{m_{j'}(W_1,\theta_n)}{\sigma_{F_n,j'}(\theta_n)}\right|^{2+\delta}\right]^{(1+\eta)/(2+\delta)} \\
&< \infty,
\end{aligned}
$$

where $\delta_3 = 2(1+\eta)(2+\delta)/(\delta - 2\eta)$, the first inequality holds by Hölder's inequality, and the second holds for sufficiently small $\eta > 0$ by Assumption M(b) and condition (vi) of (2.3).

With the manageability of $\{f^{mm}_{n,i,j,j'}(\omega, g, g^*) : g, g^* \in \mathcal{G}, i \leq n, n \geq 1\}$ and (16.26), Lemma E2 gives

$$
(16.27)\quad \sup_{g,g^*\in\mathcal{G}} \left| n^{-1}\sum_{i=1}^n f^{mm}_{n,i,j,j'}\bigl(\omega, g, g^*\bigr) - Ef^{mm}_{n,1,j,j'}\bigl(\cdot, g, g^*\bigr)\right| \to_p 0.
$$

By (16.16), (16.22), (16.27), as well as $|Ef^{mm}_{n,1,j}(\cdot, g)| \leq E(F^m_{n,1,j})^{1+\eta} < \infty$, we conclude that the difference between the (j, j')th element of $D^{-1/2}_{F_n}(\theta_n)\widehat{\Sigma}_n(\theta_n, g, g^*)D^{-1/2}_{F_n}(\theta_n)$ and $Ef^{mm}_{n,1,j,j'}(\cdot, g, g^*) - Ef^m_{n,1,j}(\cdot, g)Ef^m_{n,1,j'}(\cdot, g^*)$ converges to zero uniformly over $(g, g^*) \in \mathcal{G}^2$.

By definition,

$$
\begin{aligned}
(16.28)\quad & Ef^{mm}_{n,1,j,j'}(\cdot, g, g^*) - Ef^{m}_{n,1,j}(\cdot, g) Ef^{m}_{n,1,j'}(\cdot, g^*) \\
&= E_{F_n}\big[\sigma^{-1}_{F_n,j}(\theta_n)\sigma^{-1}_{F_n,j'}(\theta_n) m_j(W_1, \theta_n) g_j(X_1) m_{j'}(W_1, \theta_n) g^*_{j'}(X_1)\big] \\
&\quad - E_{F_n}\big[\sigma^{-1}_{F_n,j}(\theta_n) m_j(W_1, \theta_n) g_j(X_1)\big] \\
&\quad \times E_{F_n}\big[\sigma^{-1}_{F_n,j'}(\theta_n) m_{j'}(W_1, \theta_n) g^*_{j'}(X_1)\big] \\
&= \sigma^{-1}_{F_n,j}(\theta_n)\sigma^{-1}_{F_n,j'}(\theta_n)\big[\Sigma_{F_n}(\theta_n, g, g^*)\big]_{j,j'} \\
&\to \big[h_2(g, g^*)\big]_{j,j'},
\end{aligned}
$$

where the convergence holds uniformly over $(g, g^*) \in \mathcal{G}^2$ by conditions (i) and (iv) in Definition SubSeq(h_2). This completes the proof of Lemma A1(b). *Q.E.D.*

16.4. *Proof of Lemma E1*

Part (a) is proved by a similar, but simpler, argument to that given in (16.24)–(16.25).

Next, we prove part (b). Because $EF_{n,i} < \infty$ and the processes $\{\phi_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ are row-wise i.i.d., $E\mathcal{F}_n \equiv \{E\phi_{n,i}(\cdot, g) \cdot 1_n : g \in \mathcal{G}\}$ is a subset of a one-dimensional affine subspace of R^n with diameter no greater than $2EF_{n,i}$. Thus, $\alpha \odot E\mathcal{F}_n$ is a subset of a one-dimensional affine subspace of R^n with diameter no greater than $2\|\alpha\|EF_{n,i}$. By Lemma 4.1 in Pollard (1990), we have: for all $n \geq 1$,

$$
(16.29)\quad D(\xi\|\alpha \odot EF_n\|, \alpha \odot E\mathcal{F}_n) \leq 6\|\alpha\|EF_{n,i}/(\xi\|\alpha \odot EF_n\|) = 6/\xi.
$$

Because $\int_0^1 \sqrt{\log(6/\xi)}\, d\xi = 3\sqrt{\pi} < \infty$, part (b) holds.

Finally, we prove part (c). Let $\lambda^*_\phi(\xi) : (0, 1] \to R^+$ be the square-root-log integrable function such that

$$
(16.30)\quad D(\xi\|\alpha \odot F^*_n(\omega)\|, \alpha \odot \mathcal{F}^*_{n,\omega}) \leq \lambda^*_\phi(\xi) \quad \text{for} \quad 0 < \xi \leq 1,
$$

for all $\alpha \in [0, \infty)^n$, $\omega \in \Omega$, and $n \geq 1$. Let

$$
\begin{aligned}
(16.31)\quad & \mathcal{F}^*_{n,\omega} = \{\phi^*_n(\omega, g) : g \in \mathcal{G}\}, \\
& \mathcal{F}^{\text{sum}}_{n,\omega} = \{\phi_n(\omega, g) + \phi^*_n(\omega, g) : g \in \mathcal{G}\}, \\
& \mathcal{F}^+_{n,\omega} = \mathcal{F}^*_{n,\omega} \oplus \mathcal{F}_{n,\omega} \equiv \{a + b \in R^n : a \in \mathcal{F}_{n,\omega}, b \in \mathcal{F}^*_{n,\omega}\},
\end{aligned}
$$

where $\phi_n(\omega, g) = (\phi_{n,1}(\omega, g), \ldots, \phi_{n,n}(\omega, g))'$. Let

$$
(16.32)\quad F^{\text{sum}}_n(\omega) = F_n(\omega) + F^*_n(\omega).
$$

Then, for $0 < \xi \le 1$ and $\alpha \in [0, \infty)^n$,

$$
\begin{aligned}
(16.33)\quad & D\big(\xi\big\|\alpha \odot F_n^{\text{sum}}(\omega)\big\|, \alpha \odot \mathcal{F}_{n,\omega}^{\text{sum}}\big) \\
&\le D\big(\xi\big\|\alpha \odot F_n^{\text{sum}}(\omega)\big\|, \alpha \odot \mathcal{F}_{n,\omega}^{+}\big) \\
&\le D\big(\xi\big(\big\|\alpha \odot F_n(\omega)\big\| + \big\|\alpha \odot F_n^*(\omega)\big\|\big)/\sqrt{2}, \alpha \odot \mathcal{F}_{n,\omega}^{+}\big) \\
&\le D\big(\xi\big\|\alpha \odot F_n(\omega)\big\|/(2\sqrt{2}), \alpha \odot \mathcal{F}_{n,\omega}\big) \\
&\quad \times D\big(\xi\big\|\alpha \odot F_n^*(\omega)\big\|/(2\sqrt{2}), \alpha \odot \mathcal{F}_{n,\omega}^{*}\big) \\
&\le \lambda_\phi\big(\xi/(2\sqrt{2})\big)\lambda_\phi^*\big(\xi/(2\sqrt{2})\big),
\end{aligned}
$$

where $\lambda_\phi(\xi)$ denotes the packing number bounding function given in Definition E2 for the processes $\{\phi_n(\omega, g) : g \in \mathcal{G}, i \le n, n \ge 1\}$, the first inequality holds because $\mathcal{F}_{n,\omega}^{\text{sum}} \subset \mathcal{F}_{n,\omega}^{+}$, the second inequality holds because $D(x, G)$ is decreasing in x and $\|a + b\| \ge (\|a\| + \|b\|)/\sqrt{2}$ for $a, b \in [0, \infty)^n$, the third inequality holds by a stability result for packing numbers (see Pollard (1990, p. 22)), and the last inequality holds by the manageability of $\{\phi_n(\omega, g) : g \in \mathcal{G}, i \le n, n \ge 1\}$ and (16.30).

The function $\lambda_\phi(\xi/(2\sqrt{2}))\lambda_\phi^*(\xi/(2\sqrt{2}))$ is square-root-log integrable by (16.25), which completes the proof of part (c). *Q.E.D.*

16.5. *Proof of Lemma E2*

We prove convergence in probability by showing convergence in L^1. We have

$$
\begin{aligned}
(16.34)\quad & E\sup_{g\in\mathcal{G}}\left|n^{-1}\sum_{i=1}^n\big[f_{n,i}(\cdot, g) - Ef_{n,i}(\cdot, g)\big]\right| \\
&\le n^{-1}KE\left(\sum_{i=1}^n F_{n,i}^2\right)^{1/2} \\
&\le n^{-1}KE\left(\sum_{i=1}^n F_{n,i}^{1+\eta}\right)^{1/(1+\eta)} \le n^{-1}K\left(E\sum_{i=1}^n F_{n,i}^{1+\eta}\right)^{1/(1+\eta)} \\
&\le n^{-\eta/(1+\eta)}K\big(B^*\big)^{1/(1+\eta)} \to 0 \quad \text{as} \quad n \to \infty,
\end{aligned}
$$

where the first inequality holds for some constant $K < \infty$ by manageability and the maximal inequality (7.10) in Pollard (1990), the second inequality holds using $0 < \eta < 1$ by applying the inequality $\sum_{i=1}^n x_i^s \le (\sum_{i=1}^n x_i)^s$, which holds for $s \ge 1$ and $x_i \ge 0$ for $i = 1, \dots, n$, with $x_i = F_{n,i}^{1+\eta}$ and $s = 2/(1+\eta) > 0$, the third inequality holds by the concavity of the function $f(x) = x^{1/(1+\eta)}$ when $\eta > 0$, and the last inequality holds because $n^{-1}\sum_{i=1}^n EF_{n,1}^{1+\eta} \le B^*$for all $n \ge 1$. *Q.E.D.*

16.6. *Proof of Lemma E3*

For notational simplicity, we prove Lemma E3 for the sequence $\{n\}$, rather than the subsequence $\{a_n\}$. All of the arguments in this subsection go through with $\{a_n\}$ in place of $\{n\}$.

The conclusions of Lemma E3 are implied by the result of Theorem 10.6 of Pollard (1990), which relies on the following five conditions:

(i) the $\{f_{ni}(\omega, g) : g \in \mathcal{G}\}$ defined in (16.4) are manageable with respect to some envelope $F_{a,n}(\omega) = (F_{a,n,1}(\omega), \ldots, F_{a,n,n}(\omega))'$,

(ii) $\lim_{n\to\infty} Ea'\nu_{n,F_n}(\theta_n, g)\nu_{n,F_n}(\theta_n, g^*)'a = a'h_2(g, g^*)a$ for all $g, g^* \in \mathcal{G}$,

(iii) $\limsup_{n\to\infty} \sum_{i=1}^n EF_{a,n,i}^2 < \infty$,

(iv) $\sum_{i=1}^n EF_{a,n,i}^2\{F_{a,n,i} > \xi\} \to 0$ as $n \to \infty$ for each $\xi > 0$, and

(v) the limit $\rho_a(\cdot, \cdot)$ is well defined by (16.7), and for all deterministic sequences $\{g_{(n)}\}$ and $\{g^*_{(n)}\}$, if $\rho_a(g_{(n)}, g^*_{(n)}) \to 0$, then $\rho_{n,a}(g_{(n)}, g^*_{(n)}) \to 0$ as $n \to \infty$.

Now we verify the five conditions:

(i) By (16.4), we have

$$\text{(16.35)}\quad f_{n,i}(\omega, g) = \sum_{j=1}^{k} a_j\sigma^{-1}_{F_n,j}(\theta_n)n^{-1/2}\big[m_j\big(W_{n,i}(\omega), \theta_n\big)g_j\big(X_{n,i}(\omega)\big) - E_{F_n}m_j(W_i, \theta_n)g_j(X_i)\big],$$

where a_j denotes the jth element of a. By Assumption M(c), $\{g_j(X_{n,i}(\omega)) : i \le n\}$ are manageable with respect to envelopes $G(X_{n,i}(\omega))$. Therefore, by Lemma E1(a)–(c), $\{f_{n,i}(\omega, g) : i \le n\}$ is manageable with respect to envelopes $F_{a,n} = (F_{a,n,1}, \ldots, F_{a,n,n})'$ defined by

$$\text{(16.36)}\quad F_{a,n,i}(\omega) = n^{-1/2}\sum_{j=1}^{k} a_j\sigma^{-1}_{F_n,j}(\theta_n)\big[\big|m_j\big(W_{n,i}(\omega), \theta_n\big)\big|G\big(X_{ni}(\omega)\big) + E_{F_n}\big|m_j(W_i, \theta_n)\big|G(X_i)\big].$$

(ii) By (16.5), we have

$$\begin{aligned}\text{(16.37)}\quad & Ea'\nu_{n,F_n}(\theta_n, g)\nu'_{n,F_n}\big(\theta_n, g^*\big)a \\ &= E\left(\sum_{i=1}^{n} f_{n,i}(\cdot, g)\right)\left(\sum_{i=1}^{n} f_{n,i}\big(\cdot, g^*\big)\right)' = nEf_{n,1}(\cdot, g)f_{n,1}\big(\cdot, g^*\big)' \\ &= n^{-1}a'D^{-1/2}_{F_n}(\theta_n)\cdot \mathrm{Cov}_{F_n}\big(m(W_1, \theta_n, g), m\big(W_1, \theta_n, g^*\big)\big)\cdot D^{-1/2}_{F_n}(\theta_n)a \\ &= n^{-1}a'D^{-1/2}_{F_n}(\theta_n)\Sigma_{F_n}\big(\theta_n, g, g^*\big)D^{-1/2}_{F_n}(\theta_n)a,\end{aligned}$$

where the second equality holds because the data are i.i.d. and the third inequality holds by (16.4). Condition (i) in Definition SubSeq(h_2) completes the verification of condition (ii) above.

(iii) Next, we verify $\limsup_{n\to\infty}\sum_{i=1}^n EF_{a,n,i}^2 < \infty$. By the linear structure of $F_{a,n,i}$, it suffices to show that

$$\text{(16.38)}\quad \limsup_{n\to\infty} E_{F_n}\sigma_{F_n,j}^{-2}(\theta_n)\big|m_j(W_i,\theta_n)\big|^2 G^2(X_i) < \infty \quad \text{and}$$
$$\limsup_{n\to\infty} E_{F_n}\sigma_{F_n,j}^{-1}(\theta_n)\big|m_j(W_i,\theta_n)\big| G(X_i) < \infty.$$

The latter is implied by the former and the former holds by the same argument as in (16.21) with $\eta = 1$.

(iv) For B as in condition (vi) of (2.3), $\xi > 0$, and $\eta > 0$ sufficiently small,

$$\begin{aligned}
\text{(16.39)}\quad & \sum_{i=1}^n EF_{a,n,i}^2\{F_{a,n,i} > \xi\} \\
&= nEF_{a,n,i}^2\{F_{a,n,i} > \xi\} \le nEF_{a,n,i}^{2+\eta}/\xi^\eta \\
&\le \frac{2(2k)^{2+\eta}}{n^{\eta/2}\xi^\eta}\sum_{j=1}^k |a_j|^{2+\eta} E_{F_n} G^{2+\eta}(X_i)\sigma_{F_n,j}^{-2-\eta}(\theta_n)\big|m_j(W_i,\theta_n)\big|^{2+\eta} \\
&\le \frac{2(2k)^{2+\eta}}{n^{\eta/2}\xi^\eta}\sum_{j=1}^k |a_j|^{2+\eta}\big[E_{F_n} G^{\delta_4}(X_1)\big]^{(2+\eta)/\delta_4} B^{(2+\eta)/(2+\delta)} \\
&\le \frac{2(2k)^{2+\eta}B^{(2+\eta)/(2+\delta)}C^{(2+\eta)/\delta_1}}{n^{\eta/2}\xi^\eta}\sum_{j=1}^k |a_j|^{2+\eta} \to 0,
\end{aligned}$$

where the first equality holds because the data are identically distributed, the second inequality holds by Jensen's inequality using the convexity of $\psi(x) = x^{2+\eta}$, that is, $((2k)^{-1}\sum_{j=1}^k(|X_j| + E|X_j|))^{2+\eta} \le (2k)^{-1}\sum_{j=1}^k(|X_j|^{2+\eta} + (E|X_j|)^{2+\eta})$ and $(E|X_j|)^{2+\eta} \le E|X_j|^{2+\eta}$, the third inequality holds with $\delta_4 = (2+\eta)(2+\delta)/(\delta-\eta)$ by the same arguments as in (16.26), and the fourth inequality holds by Assumption M(b) and $\delta_4 \le \delta_1$ for sufficiently small η.

(v) First we show that the limit $\rho_a(\cdot,\cdot)$ is well defined by (16.7). For any $g, g^* \in \mathcal{G}$,

$$\begin{aligned}
\text{(16.40)}\quad & \rho_{n,a}^2(g,g^*) \\
&= nE\big(f_{n,i}(\cdot,g) - f_{n,i}(\cdot,g^*)\big)^2 \\
&= a'D_{F_n}^{-1/2}(\theta_n)\operatorname{Var}_{F_n}\big(m(W_i,\theta_n,g) - m(W_i,\theta_n,g^*)\big)D_{F_n}^{-1/2}(\theta_n)a \\
&\to a'h_2(g,g)a + a'h_2(g^*,g^*)a - a'h_2(g,g^*)a - a'h_2(g^*,g)a,
\end{aligned}$$

where the convergence hold uniformly over $\mathcal{G}^2$ by condition (i) in Definition SubSeq(h_2). Thus, $\rho_a(g, g^*) = \lim_{n\to\infty} \rho_{n,a}(g, g^*)$ is well defined, and

$$\lim_{n\to\infty} \sup_{g,g^*\in\mathcal{G}} \left|\rho_{n,a}(g, g^*) - \rho_a(g, g^*)\right| = 0. \tag{16.41}$$

Lastly, we show the second property of condition (v). Let $\xi > 0$ be arbitrary. Suppose $\rho_a(g_{(n)}, g^*_{(n)}) \to 0$. Then, there exists an $N_0 < \infty$ such that, for $n \geq N_0$,

$$\rho_a\left(g_{(n)}, g^*_{(n)}\right) \leq \xi/2. \tag{16.42}$$

By (16.41), we have

$$\lim_{m\to\infty} \sup_{n\geq 1}\left|\rho_{m,a}\left(g_{(n)}, g^*_{(n)}\right) - \rho_a\left(g_{(n)}, g^*_{(n)}\right)\right| = 0. \tag{16.43}$$

Thus, there exists an $N_1 < \infty$ such that, for all $m \geq N_1$,

$$\sup_{n\geq 1}\left|\rho_{m,a}\left(g_{(n)}, g^*_{(n)}\right) - \rho_a\left(g_{(n)}, g^*_{(n)}\right)\right| \leq \xi/2. \tag{16.44}$$

Take $N = \max\{N_0, N_1\}$; then we have, for $n \geq N$,

$$\rho_{n,a}\left(g_{(n)}, g^*_{(n)}\right) \leq \xi. \tag{16.45}$$

Thus, $\rho_a(g_{(n)}, g^*_{(n)}) \to 0$ implies $\rho_{n,a}(g_{(n)}, g^*_{(n)}) \to 0$. *Q.E.D.*

17. SUPPLEMENTAL APPENDIX F

In Sections 17.1 and 17.5, this appendix provides additional material concerning the Monte Carlo simulations in the quantile selection and entry game models. In Sections 17.2 and 17.4, it provides all of the Monte Carlo simulation results for the mean selection and interval-outcome regression models. In Section 17.3, it provides some results for CLR-series CI's with different upper bounds on the number of series terms considered by the cross-validation procedure that is used to select the number of series terms.

17.1. *Quantile Selection Model*

Section 17.1.1 provides additional simulation results to those given in the paper. Section 17.1.2 provides figures for the conditional moment functions evaluated at the θ values at which the FCP's are computed in Table IV of the paper. Section 17.1.3 describes the computation of the Chernozhukov, Lee, and Rosen (2013) (CLR) and Lee, Song, and Whang (2011) (LSW) CI's.

TABLE S-I
QUANTILE SELECTION MODEL, KINKED BOUND, AND LOWER ENDPOINT: VARIATIONS ON THE BASE CASE

		(a) Coverage Probabilities		(b) False CP's (CP-corrected)	
	Statistic:	CvM/Max	KS/Max	CvM/Max	KS/Max
Case	Crit Val:	GMS/Asy	GMS/Asy	GMS/Asy	GMS/Asy
Base Case ($n = 250, r_1 = 7, \varepsilon = 5/100$)		.983	.984	.34	.52
$n = 100$		.981	.985	.34	.55
$n = 500$		.984	.984	.39	.54
$n = 1000$		.984	.980	.41	.54
$r_1 = 5$		.981	.981	.34	.49
$r_1 = 9$		.983	.986	.35	.55
$r_1 = 11$		.984	.987	.36	.60
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$		.984	.997	.39	.51
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$		.990	.991	.38	.59
$\varepsilon = 1/100$		.981	.981	.34	.56
$\alpha = .5$		.721	.710	.03	.06
$\alpha = .5$ & $n = 500$		.741	.734	.04	.08

17.1.1. *Additional Simulation Results*

Table S-I reports CP and FCP results for variations on the base case for the lower endpoint with the kinked bound DGP. (Table III of AS reports analogous results for the lower endpoint with the flat bound.) The results are similar to those in Table III of AS. There is relatively little sensitivity to the sample size, the number of cubes g, and the choice of ε. There is relatively little sensitivity of the CP's to the choice of (κ_n, B_n), but some sensitivity of the FCP's with the base case choice being superior to values of (κ_n, B_n) that are twice or half as large. The CI with $\alpha = .5$ is half-median unbiased and avoids the well-known problem of inward-bias. But, it is farther from being median-unbiased than in the flat bound case.

Next, Table S-II provides coverage probability (CP) and false coverage probability (FCP) results for the upper endpoint of the identified interval in the quantile selection model.[60] (Table I of AS provides analogous results for the lower endpoint.) Table S-II provides a comparison of CI's based on the CvM/Sum, CvM/QLR, CvM/Max, KS/Sum, KS/QLR, and KS/Max statistics, coupled with the PA/Asy and GMS/Asy critical values. The relative attributes of the different CI's are quite similar to those reported in Table I of AS for the lower endpoint. None of the CI's under-cover. So, the relative attributes of

[60]For the upper endpoint with the flat bound and the upper endpoint with the kinked bound, the FCP's are computed at the points $\underline{\theta}(1) + 0.40 \times \text{sqrt}(250/n)$ and $\underline{\theta}(1) + 0.75 \times \text{sqrt}(250/n)$, respectively. These points are chosen to yield similar values for the FCP's across the different cases considered.

TABLE S-II

QUANTILE SELECTION MODEL, UPPER ENDPOINT: BASE CASE TEST STATISTIC COMPARISONS

		Statistic					
DGP	Crit Val	CvM/Sum	CvM/QLR	CvM/Max	KS/Sum	KS/QLR	KS/Max
		(a) Coverage Probabilities					
Flat Bound	PA/Asy	.994	.994	.993	.984	.984	.982
	GMS/Asy	.971	.971	.970	.974	.974	.972
Kinked Bound	PA/Asy	.996	.996	.996	.989	.989	.988
	GMS/Asy	.974	.974	.972	.976	.976	.975
		(b) False Coverage Probabilities (coverage probability corrected)					
Flat Bound	PA/Asy	.73	.72	.71	.70	.70	.69
	GMS/Asy	.42	.42	.42	.55	.55	.55
Kinked Bound	PA/Asy	.73	.73	.72	.74	.74	.73
	GMS/Asy	.41	.41	.41	.52	.52	.52

the CI's are determined by their FCP's. The CvM-based CI's have lower FCP's than the KS-based CI's. The CI's that use the GMS/Asy critical values have lower FCP's than those based on the PA/Asy critical values. The FCP's do not depend on whether the Sum, QLR, or Max version of the statistic is employed. Hence, the best CI of those considered is the CvM/Max/GMS/Asy CI, or this CI with Max replaced by Sum or QLR.

17.1.2. *Conditional Moment Function Figures*

Figure S-1 shows the conditional moment functions $\beta(x, \theta)$ (defined in (10.6)), as functions of x, evaluated at the θ values 1.531, 1.181, and 1.151 at which the FCP's are computed in Tables I and II of the paper in the flat, kinked, and peaked cases, respectively.

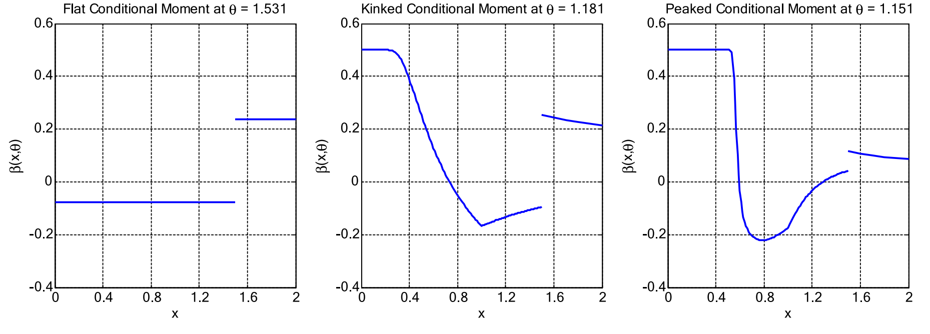

FIGURE S-1.—Conditional moment functions for the quantile selection model evaluated at θ values below the lower endpoint of the identified set.

17.1.3. *Description of the CLR-Series, CLR-Local Linear, and LSW Confidence Intervals*

Here we describe the computation of the CLR and LSW CI's reported in Table IV for the quantile selection model. These are two-sided CI's. In the quantile selection model, the parameter θ is not separable from its bound functions. In consequence, for the CLR CI's, we follow the method in Example C in the 2011 version of CLR. We define an auxiliary parameter β:

$$\beta(\theta) = \min_{x \in R} \beta(x, \theta), \quad \text{where} \tag{17.1}$$

$$\beta(x, \theta) = \begin{cases} E\big(1(Y_i \leq \theta, T_i = t) + 1(T_i \neq t) - \tau | X_i = x\big), & \text{if } x < x_0, \\ E\big(\tau - 1(Y_i \leq \theta, T_i = t) | X_i = x\big), & \text{if } x \geq x_0. \end{cases} \tag{17.2}$$

We obtain a CLR bound estimator $\hat{\beta}_\alpha(\theta)$ for a null θ value using the GAUSS code provided by CLR and let the nominal $1 - \alpha$ confidence set for θ be $CS_n^{CLR}(\alpha) = \{\theta : \hat{\beta}_\alpha(\theta) \geq 0\}$.[61] Note that the CLR CI is constructed using the auxiliary parameter β, for which the bound is one-sided. Therefore, the one-sided inference procedure of CLR is applied, even though the resulting confidence set for θ is two-sided (if it is an interval).

To implement the LSW confidence set in the quantile selection model, for each θ, we use LSW's test for the null hypothesis $H_0 : -\beta(x, \theta) \leq 0, \forall x \in \mathcal{X}$ and let the confidence set consist of all of the θ values for which the test does not reject the null. We use the GAUSS code provided by LSW to carry out the LSW test.[62]

We report results for the L^1 version of the LSW CI with inverse standard deviation weight function and bandwidth parameter $c_h = 2.0$. These choices provide the best overall performance of the LSW CI. As noted in a footnote 29 in Section 10.1 of the paper, we use the inverse standard deviation weight function, rather than the uniform weight function, because the CI that uses the latter performs very poorly in terms of FCP's in the cases reported in Table V of the paper.

As noted in footnote 28 in Section 10.1 of the paper, the number of series terms is selected by cross-validation with an upper bound of 30 on the number of series terms with the CLR-series CI, whereas CLR used an upper bound of 9 in the 2011 version of their paper. The footnote explains why. Briefly, the choice of 9 performs very poorly in terms of CP's in the cases considered in Table V of the paper. The lower bound on the number of series terms is 5, as in CLR.

[61]See the simulation section of the 2011 version of CLR for a description of what the code does. We thank CLR for making their code available to us.

[62]See the simulation section of the 2012 version of LSW for a description of what the code does. We thank LSW for making their code available to us.

17.2. *Mean Selection Model*

17.2.1. *The Model*

In this section, we consider the same mean selection model that is considered in the 2009 working paper version of CLR (which considers the CLR CI's, but not the AS and LSW CI's). As in the latter paper, all of the CI's considered with this model are one-sided CI's of the form $[\widehat{lb}_n, \infty)$ for some random variable $\widehat{lb}_n$. Hence, only a single moment inequality is considered and the Sum, Max, and QLR statistics are identical. We compare the CP's and FCP's of the CI's based on the CvM and KS statistics and the PA and GMS critical values.[63] We also compare the CvM/Max/GMS/Asy CI (abbreviated by AS below) with several other CI's in the literature, that is to say, the CLR-series, CLR-local linear, and LSW CI's.[64]

The model is essentially the same as the quantile selection model described in the paper, except that the parameter of interest θ is the conditional mean $E(y_i(1)|X_i = x_0)$ for some x_0, rather than the conditional quantile. In addition, the QMIV assumption is replaced with the monotone instrumental variable (MIV) assumption of Manski and Pepper (2000): for all $(x_1, x_2) \in \mathcal{X}^2$ such that $x_1 \leq x_2$,

$$E\big(y_i(1)|X_i = x_1\big) \leq E\big(y_i(1)|X_i = x_2\big). \tag{17.3}$$

The MIV assumption is not informative unless $y_i(t)$ has bounded support. Let the support of $y_i(1)$ be $[Y_l, Y_u]$. The MIV assumption leads to the following moment inequalities:

$$\begin{aligned} &E\big(1(X_i \leq x_0)\big[\theta - Y_i 1(T_i = 1) - Y_l 1(T_i \neq 1)\big]|X_i\big) \geq 0 \quad \text{a.s.} \quad \text{and} \\ &E\big(1(X_i \geq x_0)\big[Y_i 1(T_i = 1) + Y_u 1(T_i \neq 1) - \theta\big]|X_i\big) \geq 0 \quad \text{a.s.} \end{aligned} \tag{17.4}$$

We only use the first inequality because we consider one-sided CI's in this model.

We consider the following data generating processes (DGP's): $y_i(1) = \mu(X_i) + u_i$ and $[Y_l, Y_u] = [-1.96, 1.96]$, where $X_i \sim \text{Unif}[-2, 2]$ and $u_i \sim 1.96 \wedge ((-1.96) \vee N(0, 1))$, $T_i = 1\{L(X_i) + \varepsilon_i \geq 0\}$, where $\varepsilon_i \sim N(0, 1)$ and ε_i, u_i, and X_i are independent of each other, and $Y_i = y_i(T_i)$. Three specifications of $(\mu(x), \sigma(x), L(x))$ are considered, which yield flat, kinked, and peaked bound functions for the conditional mean θ. For the flat bound DGP, $\mu(x) = 0 = L(x)$. For the kinked bound DGP, $\mu(x) = 2(x \wedge 1)$ and $L(x) = x \wedge 1$. For the peaked bound DGP, $\mu(x) = 2|x - 1|$ and $L(x) = (x \wedge 1)$. The

[63] These comparisons are similar to those given in Table I of the paper for the quantile selection model, but all of the CI's are one-sided, not two-sided.

[64] These comparisons are similar to those given in Table IV of the paper for the quantile selection model, but all of the CI's are one-sided, not two-sided.

parameter of interest is the conditional mean of $y_i(1)$ at $x_0 = 1.5$, that is, $\theta = E(y_i(1)|X_i = 1.5)$.

17.2.2. *Description of the CLR-Series, CLR-Local Linear, and LSW Confidence Intervals*

Next, we describe the computation of the CLR and LSW CI's reported in Table S-IV (given below) for the mean selection model. In this model, all of the CI's are one-sided CI's. In the mean selection model, the parameter θ is separable from its bound function:

$$\theta \geq \sup_{x \in R} \theta_l(x), \quad \text{where} \tag{17.5}$$

$$\theta_l(x) = E\big(Y_i 1(T_i = t) + Y_l 1(Y_i \neq t)|X_i = x\big) \quad \text{for} \quad x < x_0.$$

We obtain a CLR bound estimator $\hat{\theta}_\alpha$ using the GAUSS code provided by CLR and let the nominal $1-\alpha$ confidence set for θ be $\mathrm{CS}_n^{\mathrm{CLR}}(\alpha) = \{\theta : \theta \geq \hat{\theta}_\alpha\}$. This yields a one-sided CI.

For the LSW CI, we compute a one-sided CI as well. For each θ, we use LSW's test for the null hypothesis $H_0 : (\theta_l(x) - \theta)1(x < x_0) \leq 0, \forall x \in \mathcal{X}$ and let the CI consist of all of the θ values such that the test does not reject the null. We use the GAUSS code provided by LSW to carry out the LSW test.

17.2.3. *Simulation Results*

We consider sample size $n = 250$ (which is also the base case sample size for the quantile selection model in the paper). All results concern the lower end of the identified interval for θ, which equals $-.98$, 1.372, and .530 in the flat, kinked, and peaked bound cases, respectively.[65] All results are based on (5000, 5001) coverage probability and critical value repetitions, respectively. The FCP's are CP-corrected, as described in Section 10 of the paper.[66]

Tables S-III and S-IV report the simulation results for the mean selection model.

Table S-III provides CP and FCP comparisons of the CI's based on the test statistics CvM/Max and KS/Max (which are equivalent to Sum and QLR versions of these statistics because only one moment function is considered) and the PA/Asy and GMS/Asy critical values. The CP results are similar to those for the quantile selection model given in Table I. All versions of the CI's have good CP's (i.e., CP's greater than or equal to .95). In contrast to the quantile

[65] The DGP is the same for FCP's as for CP's; just the value θ that is to be covered is different. For the lower endpoint of the identified set, FCP's are computed for θ equal to $\underline{\theta}(1) - c$, where $c = .155$, .68, and .78 in the flat, kinked, and peaked bound cases, respectively. These points are chosen to yield similar values for the FCP's across the three cases.

[66] That is, a positive constant is added to the critical value such that the CP for the given case being considered is .95 whenever the CP for the given case (without correction) is less than .95.

TABLE S-III

MEAN SELECTION MODEL: BASE CASE TEST STATISTIC AND CRITICAL VALUE COMPARISONS

DGP	Crit Val	Statistic CvM	Statistic KS
	(a) Coverage Probabilities		
Flat Bound	PA/Asy	.953	.961
	GMS/Asy	.947	.959
Kinked Bound	PA/Asy	1.000	.996
	GMS/Asy	.958	.931
Peaked Bound	PA/Asy	1.000	1.000
	GMS/Asy	.999	.998
	(b) False Coverage Probabilities (coverage probability corrected)		
Flat Bound	PA/Asy	.37	.63
	GMS/Asy	.37	.63
Kinked Bound	PA/Asy	.81	.55
	GMS/Asy	.35	.34
Peaked Bound	PA/Asy	.58	.66
	GMS/Asy	.38	.57

selection model, the CvM/Max/PA/Asy has CP close to .95 in the flat bound case. This can be attributed to the one-sided nature of the CI's in Table S-III. Those in the quantile selection model are two-sided.

The FCP results in Table S-III also are similar to those in Table I. The GMS/Asy critical value outperforms the PA/Asy critical value in the kinked and peaked bound cases and is equally good in the flat bound case. When using the GMS/Asy critical values, the CvM statistic outperforms the KS version in terms of FCP's in the flat and peaked bound cases and has equally good performance in the kinked bound case. The main differences between the FCP results in Table S-III and Table I are (i) the GMS/Asy critical value has equal performance

TABLE S-IV

MEAN SELECTION MODEL: COMPARISONS OF AS CONFIDENCE INTERVALS WITH THOSE PROPOSED IN CLR AND LSW

CI	CP (95%) Flat	CP (95%) Kink	CP (95%) Peak	FCP (corrected) Flat	FCP (corrected) Kink	FCP (corrected) Peak	CP (50%) Flat	CP (50%) Kink	CP (50%) Peak
$n = 250$									
CvM/Max/GMS/Asy	.947	.958	.999	.37	.35	.38	.46	.63	.97
CLR-series	.946	.893	.983	.77	.38	.35	.51	.57	.91
CLR-local linear	.947	.930	.987	.68	.30	.30	.49	.69	.91
LSW	.939	1.000	1.000	.35	.86	.90	.57	.95	1.00

to the PA/Asy critical value in the flat bound case, rather than better performance, and (ii) the CvM form of the test statistic has equal performance to the KS version in the kinked bound case, rather than better performance.

Overall, Table S-III shows that the CvM/Max statistic combined with the GMS/Asy critical value performs very well. It has CP equal to .947 in the flat bound case and CP greater than .95 in the kinked and peaked bound cases. It has the lowest FCP in the flat and peaked bound cases and very close to the lowest FCP in the kinked bound case.

Table S-IV compares the AS CI with the CLR-series, CLR-local linear, and LSW CI's in terms of CP's and FCP's. The AS CI has good CP properties, that is to say, CP's greater than or equal to .947. The CLR-series and CLR-local linear CI's have minimum CP's over the three bounds of .893 and .930, which demonstrates that their finite-sample sizes are less than .95, but not too much less for the local linear version (at least if the least favorable case is among the three cases considered). The LSW CI has minimum CP of .939 over the three bounds, which is close to .95. Compared to the results in Table IV for $n = 250$ (which gives results for two-sided CI's in the quantile selection model), the CP performance of AS is the same, CLR-local linear is better, CLR-series is worse, and LSW is slightly worse (because of under-coverage in the flat bound case in Table S-IV).

The LSW and AS CI's have clearly the best (CP-corrected) FCP's for the flat bound case, with the LSW CI being slightly better than the AS CI. The CLR-local linear and CLR-series CI's have best (CP-corrected) FCP's for the kinked and peaked bound cases by a relatively narrow margin over the AS CI. The LSW CI has poor FCP's in the kinked and peaked cases.

The FCP performances of the AS CI relative to the CLR CI's in Table S-IV compared to Table IV (with $n = 250$) are better in Table S-IV for the flat and peaked bound cases and a little worse for the kinked bound case. The FCP performances of the LSW CI relative to the other CI's are better in the flat bound case in Table S-IV compared to Table IV (with $n = 250$) and worse in the kinked and peaked cases in Table S-IV compared to Table IV.

17.3. *CLR-series CI's With Different Cross-Validation Upper Bounds*

In this section, we present additional simulation results for CLR-series CI's. Specifically, we report results where the upper bound for the number of series terms used in the cross-validation procedure used to determine the number of series terms is 9, which is the choice used in the 2011 version of CLR, rather than 30, which is used in Tables IV and V of the main paper. Footnote 28 in Section 10.1 of the paper provides the reason for using an upper bound of 30. Briefly, the reason is that an upper bound of 9 yields very poor performance in terms of CP's in the cases reported in Table S-IV below. For ease of comparison, results also are reported for the AS/CvM/Max/GMS/Asy CI.

Tables S-V, S-VI, and S-VII show that the CLR-series CI's are sensitive to the upper bound used in the cross-validation procedure in some cases. The CP's

TABLE S-V

QUANTILE SELECTION MODEL: COMPARISONS OF NOMINAL 95% CLR-SERIES CI'S WITH CROSS-VALIDATION UPPER BOUNDS OF 30 AND 9

	CP (95%)			FCP (corrected)			CP (50%)		
CS	Flat	Kink	Peak	Flat	Kink	Peak	Flat	Kink	Peak
$n = 100$									
CvM/Max/GMS/Asy	.957	.981	.989	.40	.34	.47	.52	.69	.73
CLR-series-30[a]	.854	.894	.862	.80	.78	.79	.51	.67	.64
CLR-series-9	.889	.954	.945	.69	.35	.19	.54	.73	.71
$n = 250$									
CvM/Max/GMS/Asy	.951	.983	.997	.37	.34	.41	.52	.72	.82
CLR-series-30	.918	.951	.937	.70	.37	.24	.55	.76	.73
CLR-series-9	.939	.972	.979	.65	.39	.18	.57	.80	.79
$n = 500$									
CvM/Max/GMS/Asy	.954	.984	.998	.36	.39	.72	.51	.74	.88
CLR-series-30	.937	.975	.978	.70	.45	.49	.57	.80	.81
CLR-series-9	.950	.987	.989	.65	.44	.33	.59	.73	.84

[a] CLR-series-30 means that the upper bound on the number of series terms used in the cross-validation procedure is 30. CLR-series-9 means that the upper bound is 9.

TABLE S-VI

PLATEAU BOUND FUNCTIONS: COMPARISONS OF NOMINAL 95% CLR-SERIES CI'S WITH CROSS-VALIDATION UPPER BOUNDS OF 30 AND 9

		CP			FCP (CP-corrected)		
		AS	CLR		AS	CLR	
	n	CvM	series-30	series-9	CvM	series-30	series-9
DGP1	100	.986	.707	.734	.84	.88	.83
	250	.975	.805	.734	.57	.82	.75
	500	.975	.872	.525	.25	.72	.66
	1000	.971	.909	.090	.03	.57	.38
DGP2	100	1.000	.394	.207	1.00	.91	.90
	250	1.000	.683	.057	1.00	.85	.87
	500	1.000	.833	.004	.97	.77	.84
	1000	1.000	.900	.000	.70	.61	.72
DGP3	100	.970	.620	.736	.70	.89	.83
	250	.969	.762	.665	.30	.83	.75
	500	.963	.854	.436	.06	.70	.66
	1000	.969	.901	.089	.00	.55	.43
DGP4	100	.998	.321	.241	.95	.91	.90
	250	.997	.612	.021	.66	.86	.89
	500	.994	.808	.000	.23	.74	.86
	1000	.994	.893	.000	.01	.59	.79

TABLE S-VII

MEAN SELECTION MODEL: COMPARISONS OF NOMINAL 95% CLR-SERIES CI'S WITH CROSS-VALIDATION UPPER BOUNDS OF 30 AND 9

	CP (95%)			FCP (corrected)			CP (50%)		
CI	Flat	Kink	Peak	Flat	Kink	Peak	Flat	Kink	Peak
$n = 250$									
CvM/Max/GMS/Asy	.947	.958	.999	.37	.35	.38	.46	.63	.97
CLR-series-30	.946	.893	.983	.77	.38	.35	.51	.57	.91
CLR-series-9	.960	.918	.992	.77	.31	.34	.52	.60	.92

and FCP's change dramatically when the upper bound is changed from 30 to 9 in the intersection bound example for all four DGP's and all four sample sizes, especially for large sample sizes; see Table S-VI. The reason is that the upper bound of 9 is binding in these cases. The CP's and FCP's also change noticeably in the quantile selection model for the smaller sample sizes, especially with $n = 100$, but also with $n = 250$; see Table S-V. This is due to the additional noise that is introduced by allowing for a greater choice in the number of series terms.

The computation times in minutes for 5000 CLR-series tests using 5001 critical value repetitions for each test (and using a 3.33 GHz processor running GAUSS 6.0) for $n = 100$, 250, and 500 are 20, 24, and 44, respectively, when the upper bound on the number of series terms is 30. The times are 10, 11, and 12, respectively, when the upper bound is 9.

17.4. *Interval-Outcome Regression Model*

17.4.1. *Description of Model*

Here we report simulation results for an interval-outcome regression model. This model has been considered by Manski and Tamer (2002, Sec. 4.5). It is a regression model where the outcome variable Y_i^* is partially observed:

$$Y_i^* = \theta_1 + X_i\theta_2 + U_i, \quad \text{where} \quad E(U_i|X_i) = 0 \quad \text{a.s.,} \qquad \text{for} \quad i = 1, \dots, n. \tag{17.6}$$

One observes X_i and an interval $[Y_{L,i}, Y_{U,i}]$ that contains Y_i^*: $Y_{L,i} = \lfloor Y_i \rfloor$ and $Y_{U,i} = \lfloor Y_i \rfloor + 1$, where $\lfloor x \rfloor$ denotes the integer part of x. Thus, $Y_i^* \in [Y_{L,i}, Y_{U,i}]$.

It is straightforward to see that the following conditional moment inequalities hold in this model:

$$\begin{aligned} &E(\theta_1 + X_i\theta_2 - Y_{L,i}|X_i) \geq 0 \quad \text{a.s.} \quad \text{and} \\ &E(Y_{U,i} - \theta_1 - X_i\theta_2|X_i) \geq 0 \quad \text{a.s.} \end{aligned} \tag{17.7}$$

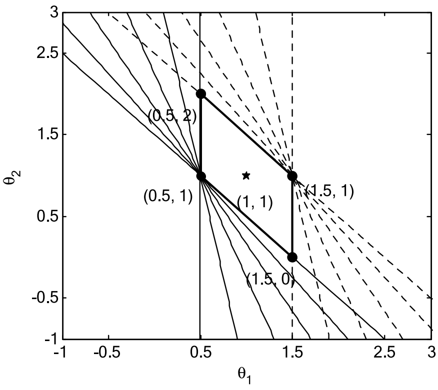

FIGURE S-2.—The identified set of the interval-outcome model.

In the simulation experiment, we take the true parameters to be $(\theta_1, \theta_2) = (1, 1)$ (without loss of generality), $X_i \sim U[0, 1]$, and $U_i \sim N(0, 1)$. We consider a base case sample size of $n = 250$, as well as $n = 100$, 500, and 1000.

The parameter $\theta = (\theta_1, \theta_2)$ is not identified. Figure S-2 shows the identified set. It is a parallelogram in (θ_1, θ_2) space enclosed by thick solid lines with vertices at $(.5, 1)$, $(.5, 2)$, $(1.5, 0)$, and $(1.5, 1)$. The point $(1, 1)$ is the true parameter. The thin solid lines are the lower bounds defined by the first moment inequality and the dashed lines are the upper bounds defined by the second moment inequality.

By symmetry, CP's of CS's are the same for the points $(.5, 1)$ and $(1.5, 1)$. Also, they are the same for $(.5, 2)$ and $(1.5, 0)$. We focus on CP's at the corner point $(.5, 1)$, which is in the identified set, and at points close to $(.5, 1)$ but outside the identified set.[67] The corner point $(.5, 1)$ is of interest because it is a point in the identified set where CP's of CS's typically are strictly less than 1. Due to the features of the model, the CP's of CS's typically equal 1 (or essentially equal 1) at interior points, non-corner boundary points, and the corner points $(.5, 2)$ and $(1.5, 0)$.

17.4.2. *g Functions*

The g functions employed by the test statistics are indicator functions of hypercubes in $[0, 1]$. It is not assumed that the researcher knows that $X_i \sim U[0, 1]$ and so the regressor X_i is transformed via the method described in

[67]Specifically, the θ values outside the identified set are given by $\theta_1 = 0.5 - 0.075 \times (500/n)^{1/2}$ and $\theta_2 = 1.0 - 0.050 \times (500/n)^{1/2}$. These θ values are selected so that the FCP's of the CS's take values in an interesting range for all values of n considered.

TABLE S-VIII

INTERVAL-OUTCOME REGRESSION MODEL: BASE CASE TEST STATISTIC COMPARISONS

	Statistic					
Critical Value	CvM/Sum	CvM/QLR	CvM/Max	KS/Sum	KS/QLR	KS/Max
			(a) Coverage Probabilities			
PA/Asy	.990	.993	.990	.989	.990	.989
GMS/Asy	.950	.950	.950	.963	.963	.963
			(b) False Coverage Probabilities (coverage probability corrected)			
PA/Asy	.62	.66	.61	.78	.80	.78
GMS/Asy	.37	.37	.37	.61	.61	.61

Section 9 to lie in $(0, 1)$.[68] The hypercubes have side-edge lengths $(2r)^{-1}$ for $r = r_0, \ldots, r_1$, where $r_0 = 1$ and the base case value of r_1 is 7. The base case number of hypercubes is 56. We also report results for $r_1 = 5$, 9, and 11, which yield 30, 90, and 132 hypercubes, respectively. With $n = 250$ and $r_1 = 7$, the expected number of observations per cube is 125, 62.5, $\ldots$, 20.8, or 17.9 depending on the cube. With $n = 250$ and $r_1 = 11$, the expected number also can equal 12.5 or 11.4. With $n = 100$ and $r_1 = 7$, the expected number is 50, 25, $\ldots$, 8.3, or 7.3.

17.4.3. *Simulation Results*

Tables S-VIII, S-IX, and S-X provide results for the interval-outcome regression model that are analogous to the results in Tables I–III for the quantile selection model. In spite of the differences in the models—the former is linear and parametric with a bivariate parameter, while the latter is nonparametric with a scalar parameter—the results are similar.

TABLE S-IX

INTERVAL-OUTCOME REGRESSION MODEL: BASE CASE CRITICAL VALUE COMPARISONS

	Critical Value				
Statistic	PA/Asy	PA/Bt	GMS/Asy	GMS/Bt	Sub
			(a) Coverage Probabilities		
CvM/Max	.990	.995	.950	.941	.963
KS/Max	.989	.999	.963	.953	.890
			(b) False Coverage Probabilities (coverage probability corrected)		
CvM/Max	.61	.69	.37	.38	.45
KS/Max	.78	.96	.61	.54	.66

[68]This method takes the transformed regressor to be $\Phi((X_i - \bar{X}_n)/\sigma_{X,n})$, where $\bar{X}_n$ and $\sigma_{X,n}$ are the sample mean and standard deviations of X_i and $\Phi(\cdot)$ is the standard normal distribution function.

TABLE S-X

INTERVAL-OUTCOME REGRESSION MODEL: VARIATIONS ON THE BASE CASE

		(a) Coverage Probabilities		(b) False Cov Probs (CPcor)	
	Statistic:	CvM/Max	KS/Max	CvM/Max	KS/Max
Case	Crit Val:	GMS/Asy	GMS/Asy	GMS/Asy	GMS/Asy
Base Case ($n=250, r_1=7, \varepsilon=5/100$)		.950	.963	.37	.61
$n=100$		.949	.970	.39	.66
$n=500$		.950	.956	.37	.60
$n=1000$		.954	.955	.37	.60
$r_1=5$ (30 cubes)		.949	.961	.37	.59
$r_1=9$ (90 cubes)		.951	.965	.37	.63
$r_1=11$ (132 cubes)		.950	.968	.38	.64
$(\kappa_n, B_n)=1/2(\kappa_{n,bc}, B_{n,bc})$		.944	.961	.40	.62
$(\kappa_n, B_n)=2(\kappa_{n,bc}, B_{n,bc})$		.958	.973	.39	.65
$\varepsilon=1/100$		.946	.966	.39	.69
$(\theta_1, \theta_2)=(1.0, 0.5)$		.999	.996	.91	.92
$(\theta_1, \theta_2)=(1.5, 0.0)$		1.000	.996	.99	.97
$\alpha=.5$		.472	.481	.03	.08
$\alpha=.5$ & $n=500$		.478	.500	.03	.07

Table S-VIII shows that the CvM/Max statistic combined with the GMS/Asy critical value has CP's that are very close to the nominal level .95. Its FCP's are noticeably lower than those for CS's that use the KS form or PA-based critical values. The CvM/Sum-GMS/Asy and CvM/QLR-GMS/Asy CS's perform equally well as the Max version. Table S-IX shows that the results for the Asy and Bt versions of the critical values are quite similar for the CvM/Max-GMS CS, which is the best CS. The Sub critical value yields substantial undercoverage for the KS/Max statistic. The Sub critical values are dominated by the GMS critical values in terms of FCP's.

Table S-X shows that the CS's do not exhibit much sensitivity to the sample size or the number of cubes employed. It also shows that at the non-corner boundary point $\theta=(1.0, 0.5)$ and the corner point $\theta=(1.5, 0)$, all CP's are (essentially) equal to 1.[69] Lastly, Table S-X shows that the lower endpoint estimator based on the CvM/Max-GMS/Asy CS with $\alpha=.5$ is close to being median-unbiased, as in the quantile selection model. It is less than the lower bound with probability .472 and exceeds it with probability .528 when $n=250$.

We conclude that the preferred CS for this model is of the CvM form, combined with the Max, Sum, or QLR function, and uses a GMS critical value, either Asy or Bt.

[69]This is due to the fact that the CP's at these points are linked to the CP's at the corner point $\theta=(0.5, 1.0)$ given the linear structure of the model. If the CP is reduced at the two former points (by reducing the critical value), the CP at the latter point is very much reduced and the CS does not have the desired size.

17.5. *Entry Game Model*

17.5.1. *Probit Log Likelihood Function*

In the entry game model, the probit log likelihood function for $\tau = (\tau_1, \tau_2)$ given $\theta = (\theta_1, \theta_2)$ is

$$
\begin{aligned}
(17.8)\qquad & \sum_{i=1}^{n} 1\big(Y_i = (0,0)\big) \ln\big(\Phi\big(-X_{i,1}'\tau_1\big)\Phi\big(-X_{i,2}'\tau_2\big)\big) \\
&\quad + \sum_{i=1}^{n} 1\big(Y_i = (1,1)\big) \ln\big(\Phi\big(X_{i,1}'\tau_1 - \theta_1\big)\Phi\big(X_{i,2}'\tau_2 - \theta_2\big)\big) \\
&\quad + \sum_{i=1}^{n} 1\big(Y_i = (1,0) \text{ or } Y_i = (0,1)\big) \ln\big(g_i(\tau, \theta)\big), \quad \text{where} \\
& g_i(\tau,\theta) = 1 - \Phi\big(-X_{i,1}'\tau_1\big)\Phi\big(-X_{i,2}'\tau_2\big) - \Phi\big(X_{i,1}'\tau_1 - \theta_1\big)\Phi\big(X_{i,2}'\tau_2 - \theta_2\big)
\end{aligned}
$$

over $\tau \in R^8$ for fixed θ. The estimator $\widehat{\tau}_n(\theta)$ maximizes this function over $\tau \in R^8$ given θ.

The gradient of the probit log likelihood for τ given θ is

$$
\begin{aligned}
(17.9)\qquad & -\sum_{i=1}^{n} 1\big(Y_i = (0,0)\big) \begin{pmatrix} \psi\big(-X_{i,1}'\tau_1\big)X_{i,1} \\ \psi\big(-X_{i,2}'\tau_2\big)X_{i,2} \end{pmatrix} \\
&\quad + \sum_{i=1}^{n} 1\big(Y_i = (1,1)\big) \begin{pmatrix} \psi\big(X_{i,1}'\tau_1 - \theta_1\big)X_{i,1} \\ \psi\big(X_{i,2}'\tau_2 - \theta_2\big)X_{i,2} \end{pmatrix} \\
&\quad + \sum_{i=1}^{n} 1\big(Y_i = (1,0) \text{ or } Y_i = (0,1)\big) \frac{1}{g_i(\tau,\theta)} \\
&\quad \times \begin{pmatrix} \phi\big(-X_{i,1}'\tau_1\big)\Phi\big(-X_{i,2}'\tau_2\big)X_{i,1} - \phi\big(X_{i,1}'\tau_1 - \theta_1\big)\Phi\big(X_{i,2}'\tau_2 - \theta_2\big)X_{i,1} \\ \Phi\big(-X_{i,1}'\tau_1\big)\phi\big(-X_{i,2}'\tau_2\big)X_{i,2} - \Phi\big(X_{i,1}'\tau_1 - \theta_1\big)\phi\big(X_{i,2}'\tau_2 - \theta_2\big)X_{i,2} \end{pmatrix},
\end{aligned}
$$

where $\psi(x) = \phi(x)/\Phi(x)$.

17.5.2. *Identification*

Here we briefly discuss identification of the entry game model. Tamer (2003, Thm. 1) provided identification results that cover the model considered in Section 10.4 because $X_{i,1}$ and $X_{i,2}$ both contain continuous regressors whose support is R.

We point out here that this support condition is probably much stronger than is needed for identification in many contexts. For example, suppose the unobservables $U_{i,1}$ and $U_{i,2}$ are independent and standard normal, as in Section 10.4. Suppose the regressor vectors are $X_{i,1} = (1, Z_i)'$ and $X_{i,2} = 1$ and their coefficient vectors are $\tau_1 = (\tau_{11}, \tau_{12})'$ and τ_2, respectively. Then, τ_1 and τ_2 are identified provided Z_i has a density with respect to Lebesgue measure on some nondegenerate interval and $\tau_{12} \neq 0$. Thus, in this case, no large support condition is needed.

To prove this result, note that $P(Y_i = (0,0)|X_{i,1}) = \Phi(-X'_{i,1}\tau_1)\Phi(-\tau_2)$. Thus, for identification at (τ_1, τ_2), it suffices to show that

$$P\big(\Phi\big(-X'_{i,1}\tau_1\big)\Phi(-\tau_2) = \Phi\big(-X'_{i,1}\lambda_1\big)\Phi(-\lambda_2)\big) = 1 \tag{17.10}$$

only if $\lambda_1 = \tau_1$ and $\lambda_2 = \tau_2$.

Suppose $\lambda_2 = \tau_2$. Then, (17.10) holds iff $P(X'_{i,1}\tau_1 = X'_{i,1}\lambda_1) = 1$. The left-hand side equals $P(\tau_{11} - \lambda_{11} + Z_i(\tau_{12} - \lambda_{12}) = 0)$. Given the condition on Z_i, the latter equals 1 only if $\lambda_1 = \tau_1$. Hence, when $\lambda_2 = \tau_2$, (λ_1, λ_2) is observationally equivalent to (τ_1, τ_2) only if $(\lambda_1, \lambda_2) = (\tau_1, \tau_2)$.

Next, suppose $\lambda_2 \neq \tau_2$. Let $c = \Phi(-\lambda_2)/\Phi(-\tau_2)$ $(\neq 1)$. Then, (17.10) holds iff $P(\Phi(-\tau_{11} - Z_i\tau_{12}) = \Phi(-\lambda_{11} - Z_i\lambda_{12})c) = 1$. The latter implies that, for all z in an open interval, say I, $\Phi(-\tau_{11} - z\tau_{12}) = \Phi(-\lambda_{11} - z\lambda_{12})c$. Taking the derivative with respect to z for $z \in I$, one obtains $\phi(-\tau_{11} - z\tau_{12}) = \phi(-\lambda_{11} - z\lambda_{12})c\lambda_{12}/\tau_{12}$. Taking logs yields a quadratic equation in z for $z \in I$:

$$\begin{aligned} &(\tau_{11} + z\tau_{12})^2 = (\lambda_{11} + z\lambda_{12})^2 + c_1 \quad \text{or} \\ &\big(\tau_{12}^2 - \lambda_{12}^2\big)z^2 + 2(\tau_{11}\tau_{12} - \lambda_{11}\lambda_{12})z + \tau_{11}^2 - \lambda_{11}^2 - c_1 = 0, \end{aligned} \tag{17.11}$$

where $c_1 = \log(c\lambda_{12}/\tau_{12})$ and c_1 is well-defined because $\tau_{12} \neq 0$. A quadratic equation cannot hold for all $z \in I$ unless each coefficient of the equation is zero, because a nondegenerate quadratic equation has at most two solutions. Suppose $\tau_{12}^2 - \lambda_{12}^2 = 0$. Then, $\tau_{11}\tau_{12} - \lambda_{11}\lambda_{12} = 0$ requires $\tau_{11} = \pm\lambda_{11}$, which implies that $\tau_{11}^2 - \lambda_{11}^2 = 0$. In consequence, $\tau_{11}^2 - \lambda_{11}^2 - c_1 = -c_1 \neq 0$ and the quadratic equation is not degenerate. (Note that $c_1 \neq 0$ because $c_1 = 0$ iff $c\lambda_{12}/\tau_{12} = 1$ iff $\lambda_{12} = c\tau_{12}$, and the latter condition violates $\tau_{12}^2 - \lambda_{12}^2 = 0$.) In conclusion, if $\lambda_2 \neq \tau_2$, (17.10) cannot hold for any λ_1 and τ_1. This completes the proof of identification.

Note that it is not clear that even continuity of Z_i in a nondegenerate interval is necessary for identification of τ. If Z_i is discrete with $s \geq 3$ support points, then observational equivalence requires s nonlinear equations in two unknowns to hold. These equations depend on the joint distribution $F(\cdot,\cdot)$ of $(U_{i,1}, U_{i,2})$. This suggests (but does not prove) that, for most joint distribution functions $F(\cdot,\cdot)$ of $(U_{i,1}, U_{i,2})$, identification holds under quite weak conditions on the regressor Z_i.

REFERENCES

ANDREWS, D. W. K., AND P. GUGGENBERGER (2009): "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709. [29,46]

——— (2010): "Asymptotic Size and a Problem With Subsampling and With the m Out of n Bootstrap," *Econometric Theory*, 26, 669–709. [39,40,64]

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157. [31]

BILLINGSLEY, P. (1995): *Probability and Measure* (Third Ed.). New York: Wiley. [34,36]

CHERNOZHUKOV, V., S. LEE, AND A. ROSEN (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737. [78]

DAVYDOV, Y. A., M. A. LIFSHITS, AND N. V. SMORODINA (1995): *Local Properties of Distributions of Stochastic Functionals*. Translations of Mathematical Monographs, Vol. 173. Providence, RI: American Mathematical Society. [22,23]

LEE, S., K. SONG, AND Y.-J. WHANG (2011): "Testing Functional Inequalities," CEMMAP Working Paper CWP12/11, Institute for Fiscal Studies, University College London. [78]

MANSKI, C. F., AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010. [82]

MANSKI, C. F., AND E. TAMER (2002): "Inference on Regression With Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546. [87]

POLITIS, D. N., AND J. P. ROMANO (1994): "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *The Annals of Statistics*, 22, 2031–2050. [64]

POLLARD, D. (1990): *Empirical Process Theory and Application*. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. II. Hayward, CA: Institute of Mathematical Statistics. [4,38,57,62,66-69,73-76]

SCHUMAKER, L. L. (2007): *Spline Functions: Basic Theory* (Third Ed.). Cambridge: Cambridge University Press. [18]

TAMER, E. (2003): "Incomplete Simultaneous Discrete Response Model With Multiple Equilibria," *Review of Economic Studies*, 70, 147–165. [91]

Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520, U.S.A.; donald.andrews@yale.edu

and

Dept. of Economics, University of Wisconsin, Madison, Madison, WI 53706, U.S.A.; xshi@ssc.wisc.edu

Manuscript received June, 2010; final revision received August, 2012.